The role of trust in proactive conversational assistants

Matthias Kraus, Nicolas Wagner, Zoraida Callejas, Wolfgang Minker

The Role of Trust in Proactive Conversational Assistants

MATTHIAS KRAUS[1], **NICOLAS WAGNER**[1], **ZORAIDA CALLEJAS**[2], **(Member, IEEE), AND WOLFGANG MINKER**[1]

[1]Institute of Communication Engineering, Ulm University, 89081 Ulm, Germany
[2]Department of Languages and Computer Systems, University of Granada, 18011 Granada, Spain

Corresponding author: Matthias Kraus (matthias.kraus@uni-ulm.de)

ABSTRACT Humans and machines harmoniously collaborating and benefiting from each other is a long lasting dream for researchers in robotics and artificial intelligence. An important feature of efficient and rewarding cooperation is the ability to assume possible problematic situations and act in advance to prevent negative outcomes. This concept of assistance is known under the term proactivity. In this article, we investigate the development and implementation of proactive dialogues for fostering a trustworthy human-computer relationship and providing adequate and timely assistance. Here, we make several contributions. A formalisation of proactive dialogue in conversational assistants is provided. The formalisation forms a framework for integrating proactive dialogue in conversational applications. Additionally, we present a study showing the relations between proactive dialogue actions and several aspects of the perceived trustworthiness of a system as well as effects on the user experience. The results of the experiments provide significant contributions to the line of proactive dialogue research. Particularly, we provide insights on the effects of proactive dialogue on the human-computer trust relationship and dependencies between proactive dialogue and user specific and situational characteristics.

INDEX TERMS Human-computer-interaction, human-computer trust, proactivity, spoken dialogue system.

I. INTRODUCTION

2021 marks the tenth anniversary of the integration of the personal assistant Siri in Apple's iPhone, introducing conversational user interfaces to the mainstream and paving the way for the nowadays ubiquitous smart assistant speakers, like Amazon Alexa or Google Echo. Being able to recognise and understand user intents, such devices are able to engage in a natural language dialogue for solving tasks cooperatively with the user. Cooperatively in this context implies that user and system take turns during an interaction exchanging information for ultimately satisfying a specific user goal, e.g. booking a restaurant, making a purchase, or simply asking for the weather. Even though being a commercial success, the assistive functions of current conversational user

The associate editor coordinating the review of this manuscript and approving it for publication was Khoa Luu.

interfaces are rather limited. For example, interactions with Alexa are mostly based on one-shot interactions (''Alexa, what's the weather today?''), only allowing a few follow-up questions, and restricted to simple task domains. However, due to technological advancement and an ever growing market for digital assistants [1], it can be expected that conversational assistants (CA) enter more sophisticated domains and be used for very challenging tasks, like decision-making [2], learning [3], or planning tasks [4]. In order to be accepted and trusted in these delicate domains, conversational interfaces must extend their assistance capabilities and be equipped with more human-like assistant behaviour.

An important aspect of future dialogue assistants will be the ability to act autonomously, hence to *proactively* provide support. In doing so, a dialogue system would make assumptions about the current situation-specific user

needs and subsequently take the floor to provide suggestions or completely act on behalf of the user. For example, an assistant that helps a user solving a particular task, e.g. planning of a city trip or working on a do-it-yourself (DIY) project, could automatically suggest possible helpful information for the user during task execution. In case a system is very sure about the user's intention it could even decide to provide the information without prompting the user.

However, there exist no clear definitions of proactive dialogue or human-computer interaction (HCI) in general. Most research applies the term proactivity to all-kind of system-initiated behaviour and focus only on the capability of machines to act autonomously [5], [6]. Fine-grained aspects of proactive behaviour, such as when to use which kind of proactive strategy or whether proactivity is necessary at all are highly understudied topics [7]. Besides, most research on proactive dialogue does not take into account the current situation and the user's characteristics. Therefore, contemporary proactive applications lack flexibility and naturalness. Additionally, initiating an interaction at an inappropriate point of time or in the wrong way could be perceived as disruptive and obtrusive. This has been shown to corrupt the human-computer relationship, especially regarding the user's perceived trust in the system [8], [9]. In summary, naive ways of proactive interaction are highly fraught with risk as misused proactivity could lead to distrust in the system and drive the user away from the system as seen by Microsoft's Office Assistant Clippit. The early proactive assistant interrupted users at inappropriate moments during task execution, providing non-helpful assistance, while behaving highly obtrusively [10]. However, we assume that proactive interaction strategies could greatly benefit the user if applied correctly. First results suggest that proactive interaction affects the user's perception of conversational systems and performs well regarding task efficiency and user satisfaction [2], [6], [11].

In order to expand this line of research, we make several contributions. First, a formalisation of proactive dialogue in CAs is provided. The formalisation is domain-independent and generalisable, which forms a framework for integrating proactive dialogue in conversational applications. Secondly, we present a study showing the relations between proactive dialogue actions and several aspects of the human-computer trust (HCT) relationship as well as effects on user experience. This article forms an extension of previous research published in Kraus *et al.* [12]. In comparison to the previous publication, we provide an extended review proactive HCI and trust in HCI. Furthermore, we integrate a previously developed set of proactive dialogue actions (None, Notification, Suggestion, Intervention) into a formalisation of proactive dialogue in CAs. Additionally, the description of the study presented in Kraus *et al.* [12] is extended by a more thorough system description and by providing new results regarding the effects of the proactive dialogue strategies. Here, we include new insights on effects on the trust

progression and relations between proactive behaviour and user characteristics.

The remaining structure of the article is as follows: Related work in the field of proactive HCI, and trust in HCI is presented in Section II. A formalisation of proactive dialogue and the definition of proactive actions is described in Section III. Subsequently, in Section IV, the development and implementation of a proactive CA for task-planning is described. The developed CA is embedded in an experimental setup for testing the relations between proactive dialogue strategies and their effects on the user. This is described in Section V. The results of the experiments are presented in Section VI. The findings from these relations between proactive dialogue and its effects on the user, are discussed in detail in Section VII. Finally, this article is concluded with a summary and an brief outlook on future work in Section VIII.

II. RELATED WORK
A. PROACTIVE HUMAN-COMPUTER INTERACTION
Cohen [13] describes the process of collaboration in communication in the following way: ''people infer the reasons behind their interlocutor's utterances and attempt to ensure their success by (at least) telling them what they need to know to be successful, and by potentially volunteering to perform actions on their behalf''. Although not mentioning explicitly the term proactivity, Cohen describes two key characteristics of proactive behaviour, actively informing the user to allow them being successful, and performing helpful actions on behalf of the user.

According to its definition in organizational psychology and management, proactivity concerns taking action in order to anticipate problematic situations and prevent negative experiences [14]–[16]. Proactive behaviour can be described by several characteristics. For example, proactivity is supposed to have a long-term focus and intends to predict future states, is action-oriented and goal-directed, while also being persistent and self-starting [14], [15]. In contrast, reactive behaviour is about reacting to environmental demands, only doing what one is told and not about developing plans to deal with possible difficulties [15]. Experiments in this research field have shown the positive effects of proactive behaviour at work. Proactivity leads to a higher job performance and team performance and is associated with leadership and innovation [14]. Additionally, it refines one's intrinsic motivation and self-regulation [15] while also creating coworker trust in work environments and positively contributing to socialisation [16]. Due to the positive effects of proactive behaviour, several researchers have studied the development of proactive human-machine interaction over the last decades [2], [7], [11], [12], [17]–[19].

One of the main streams for research on proactive behaviour forms human-robot interaction. Here, proactivity of a robot can be differentiated into three categories: approaching a human in an acceptable manner [20]–[22], sharing tasks between robot and human based on the user's

TABLE 1. Levels of autonomy according to Sheridan and Verplank [29].

Level of Autonomy	Description
1.	The computer offers no assistance; the human must take all decisions and actions.
2.	The computer offers a complete set of decision/action alternatives.
3.	The computer narrows the selection down to a few.
4.	The computer suggests one alternative.
5.	The computer executes that suggestion if the human operator approves
6.	The computer allows the human a restricted time to veto before automatic execution
7.	The computer executes automatically, then necessarily informs the human
8.	The computer informs the human only if asked
9.	The computer informs the human only if it, the computer, decides to.
10.	The computer acts completely autonomously

intention [23], [24], and assisting proactively [2], [11], [25]. With regard to the work presented in this article, it is particularly interesting to consider related work on proactive robot assistance. Application domains for robotic assistants span various fields, such as shopping assistants [2], [26], caregivers [25], or DIY-assistants [12], where robots proactively engage in a conversation for providing extra information, prompting messages or proposing actions. Related work often only studies one form of proactivity as opposed to reactive behaviour [6], [11], [26], [27]. More interesting with regard to the research presented in this article, are the effects of the robot's level of autonomy on the interaction [2], [28]. According to Sheridan and Verplank's [29] seminal work, a system/robot's autonomy level can be divided into ten levels, ranging from offering no assistance to completely autonomous behaviour. Parasurman *et al.* [30] later refined this model for HCI. The individual levels are listed in Table 1. Based on these frameworks, Peng *et al.* [2], for example, designed the proactivity of a robot in three dimensions - low, medium, and high - where the lowest level of proactivity can be considered as reactive behaviour. Depending on the proactive level the robot makes assumptions about the user's needs and either lets the user verify them (medium-level) or directly takes action (high-level). Besides *how* the robot should behave proactively, i.e. decide on the level of proactivity, also the timing *when* to be proactive has been studied for robotic assistants. For example, Grosinger *et al.* [25] equipped a robot with planning capabilities and took into account context- and time-related measures to decide when to take action. Here, the robot could infer the human's activity using a simple user model containing the user's location and day time. Depending on the activity the robot would proactively prompt the user. In the work of Liu *et al.* [26], a robot learns the appropriate moments for being proactive from using human interaction data as training input. In this context, proactive robot behaviour is triggered in moments of silence or after users' provided backchannel utterances,

e.g., "Okay", or "I see" depending on previous interaction context.

Transferring the concept of proactivity to HCI in general, the concept is mostly applied to personal assistants supporting the user in accomplishing complex tasks [31]–[34]. Here, proactive behaviour aims to initiate helping actions to avoid problems in advance. In this regard, Nothdurft *et al.* [7] describe proactivity in dialogue assistants as "an autonomous, anticipatory system-initiated behaviour, with the purpose to act in advance of a future situation, rather than only reacting to it". Under this consideration, several different dialogue actions in personal assistants can be connoted as proactive behaviour. In [35], proactivity is described as the act of leading the dialogue and to actively changing the discussion topic, while keeping the dialogue natural, coherent and engaging. Similarly, [5] understand the act of actively presenting or recommending topics related to the current interaction as proactive behaviour. Here, proactivity is especially used for resolving ambiguous user demands, by showing possible candidates for the unclear query instead of doing nothing. In general, conversational recommendation (e.g. see [36]) can be seen as proactive interaction, as it facilitates the user's item selection in large decision spaces.

For task-oriented dialogue systems, [18] introduced proactive units at system turns. These units augment system responses to user requests and contain assumable relevant information that has not yet been provided by the user. For example, by not only providing a user requested departure time in a train schedule scenario, but additionally its arrival time. While presented literature considers proactive dialogue as providing additional information, recommending information, or simply leading the interaction, this work considers proactive dialogue in the sense of mixed-initiative user interface design [17].

In mixed-initiative user interactions, a user and an autonomous assistant collaborate for solving tasks. For providing assistance, the agent needs to track the user's activities and goals while reasoning about the costs and benefits of taking automated actions. Here, proactive dialogue serves for communicating and negotiating a system's decision process for minimizing the risk of system failure. Depending on the user and the context, it may be sometimes better to not interrupt the user, while there may be situations in which suggestions could be beneficial. For instance, the proactive assistant CALO [32] makes use of reasoning for estimating the cost-benefit value of proactive behaviour and adjusts its proactivity level accordingly. The cost-benefit value is calculated using system-related metrics such as time-sensitivity of the suggestion, the degree of uncertainty and the system's confidence. The decision which level to choose is based on hand-crafted thresholds.

The way in which a proactive assistant can cooperate in mixed-initiative interaction moves along the Interface-Proactivity (IP) continuum introduced by Isbell and Pierce [33]. They transferred the levels of autonomy

introduced by Sheridan and Verplank [29] to the domain of HCI, resulting in five different levels of proactive assistant behaviour. The IP continuum ranges from zero, i.e., the user acts fully on their own, to full automation, i.e., the assistant acts fully on behalf of the user. The nuances between these two extremes form alerts, telling the user to pay attention, notifications, telling the user exactly what to pay attention to, and suggestions, providing the user with several decision options. The more proactive a system becomes, the more it takes off control and responsibilities from the user. Hence, the risk of failure also increases, as the possibility that the system might take actions incongruently with the user's goal without asking for confirmation expand, hurting the human-computer relationship [33]. Therefore, most current applications in this area deal with notification or suggestion management as the cost-benefit relation is more controllable [31]. For example, there exists a body of work on desktop [34], smartphone notifications [37], as well as proactive suggestions for interactive television [38]. Challenges for the design of proactive strategies in CAs were stated by Nothdurft *et al.* [7]. According to the authors, it is necessary to investigate *when* and *how* proactivity should be initiated for providing a positive user experience. In that sense, Yorke-Smith [32] proposed guidelines for the design of proactive behaviour in intelligent assistants. A proactive agent should be valuable to the user as it advances his or her interests and tasks. It should be aware of the current situation and act according to its abilities and knowledge. Moreover, user's should be in control of the assistant and be able to understand its actions. The agent should act unimposing and not interfere with the user's own activities and attention. For adding value to the user, a proactive assistant needs to be aware of current and future needs and opportunities and act in a safe way, minimising the risk of negative outcomes in the opinion of the user.

Following these guidelines, we present concrete steps on how to model proactive dialogue in an intelligent CA and address the challenges of proactive behaviour stated by [7]. Therefore, we formalise a proactive dialogue model and introduce four proactive dialogue actions that were derived from the presented levels of autonomy: *None*, *Notification*, *Suggestion*, *Intervention*. For the development of proactive dialogue strategies, we studied the effects of the proactive dialogue actions as well as the timing of the actions on the user experience. In proactive dialogue, a high degree of collaboration for achieving mutual goals is necessary, making humans vulnerable to the decisions of their assistant. Hence, a particular focus for the evaluation is set on the perception of the proactive system's trustworthiness.

B. TRUST IN HUMAN-COMPUTER INTERACTION

Trust forms a fundamental concept in interpersonal relationships [40]–[42], organisational management [43], [44], and human-automation interaction [45]–[50]. According to the definition provided by Mayer *et al.* [43] trust can be specified as the willingness to take risks and to be vulnerable to the

actions of another party "based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party". Due to advancements in automation technology and industry 4.0 (e.g. see [51]), collaboration with intelligent computers and machines increases and presents a shift of control towards the automated assistance. This makes a formation of trust in the user indispensable, otherwise it possibly will not be accepted and becomes obsolete. Schaefer and Hancock [47] describe the effect as the "no trust – no use" principle. This is also stated in the earlier works on trust in automation by Muir *et al.* [48], [49], who hypothesise that independently of the "intelligence" or finesse of an autonomous system, users will reject a system when it is not perceived trustworthy. In literature, this phenomena is known under the term "under reliance" [45], [50]. An example for this is the false alarm problem often occurring with fire detectors [45]. In case the false alarm ratio is too high, people may disuse the device, even though this could have negative consequences. Contrarily, "over reliance" in automation may lead to misuse because people may overestimate the competence of a system [45]. Therefore, a calibration of trust is necessary, in which a user sets an appropriate trust level corresponding to the machine's trustworthiness and uses it in accordance with its abilities and limits [48].

Trust in automated technology can be generally defined as "the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability" [46, p.51]. As indicated by the definition, three factors seem to be fundamental for modelling trust: the human, the autonomous partner, and the environment. Each factor has specific characteristics that influence the human-automation trust relationship. An extensive review hereby can be found in Schaefer and Hancock [47]. Including these factors, Hoff and Bashir [50] presented a three-layered model of trust: Dispositional trust represents a user's long term tendency to an autonomous system dependent on individual characteristics. The other two layers, situational and learned trust are controlled by the users past experience either with the environment, i.e. task type and context but also a user's self-confidence or mood, or specific features of the autonomous systems. Taking into account the dynamic nature of trust on a more short-term level, the authors distinguish between initially learned trust depending on preexisting knowledge and dynamically learned trust that is possible to change during an interaction being subject to system performance and design.

A higher degree of automation in human-machine collaboration also requires more sophisticated interaction strategies from a system's perspective in order to enhance trust and to avoid miscommunication. Therefore, intelligent automation needs be able to conduct dialogues at appropriate moments in order to be perceived trustworthy. Thus, conversational assistance is supposed to be an important factor of future automation. Many concepts of trust in automation are supposed to be easily transferable to the domain of CAs,
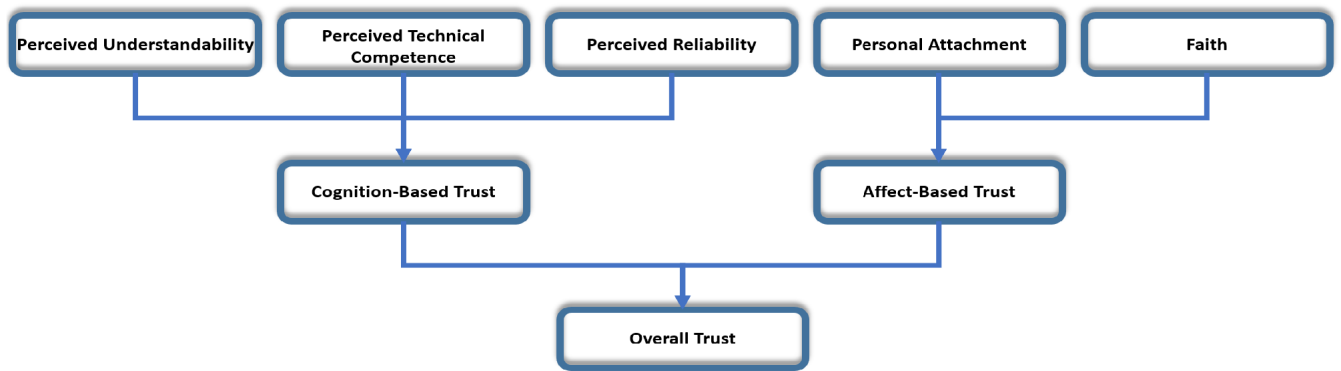
FIGURE 1. According to Madsen and Gregor [39], human-computer trust is based on the two foundations cognitive- and affect-based trust. Each base of trust comprises several sub-concepts.

as they form a specialised form of automation themselves. However, due to the ability to conduct natural dialogue and interacting on a more complex information level, some idiosyncrasies need to be taken into consideration. For example, a dialogue can have a social, e.g., small talk [52], or utilitarian purpose, e.g. solving a cooperative task. Although there exists previous work studying the effects of socio-emotional dialogue, small talk, empathetic reactions, and voice characteristics of a conversational agent on the user's perceived trust (see Rheu *et al.* [53] for an overview), most users rather see conversational agents still as ''tools'' and use the term trust with reference to a system's performance or privacy [54].

Therefore, trust in CAs resembles more the concept of computer believability or credibility as presented by Fogg and Tseng [55]. Here, a system's trustworthiness and expertise form the bases for its credibility, whereby under the term trustworthiness the quality of information (unbiased, truthful, honest) is understood. However, besides the content an agent provides, also its behaviour greatly influences its relationship with the user. Trust in the system's behaviour in utilitarian terms is mostly related to its performance with regard to consistency and reliability [46]. In this context, much research focuses on a system's capability to conduct explanation dialogues for mitigating the effects of system failures by providing transparency which in turn increases trust (e.g see Nothdurft *et al.* [56] or Glass *et al.* [57]).

Measuring trust in CAs is complicated because trust is multi-faceted and also a latent variable that cannot be observed directly. For this reason, several approaches for assessing the HCT relationship have been proposed. Primarily, subjective measurements in the form of self-reported questionnaires are collected [39], [58]. For measuring the trustworthiness of the developed strategies, the Trust in Automated Systems scale [59] and some variants [60] were used, where subjects could agree or disagree with statements about the system's impression. Sub-components of trust were measured using the HCT-model by Madsen and Gregor [39]. The model is visualised in Fig. 1. This hierarchical model relates to five fundamental components of

trust: Personal attachment and faith form the bases for affect-based trust while perceived understandability, perceived technical competence, and perceived reliability are 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. In contrast, cognition-based trust refers to a more short-termed trust. For the latter, mostly the functionality and usability of a system are of importance.

Considering trust could be beneficial for developing proactive dialogue strategies, as the degree of proactive behaviour of a technical system has shown to correlate with the HCT relationship. For example, Rau *et al.* [11] compared two levels of autonomy, high versus low. The authors presented a WoZ-study, in which participants had to complete a sea survival task in collaboration with a remotely controlled robot. The study results demonstrated that trust in the robot was higher in the low-level (reactive) than in the high-level condition.

Therefore, the effects of proactive dialogue strategies on the HCT relationship are studied in this article. Proactive dialogue forms a new way to integrate the user in the decision processes of an automated assistant in a mixed-initiative interaction. For equipping a future CA with proactive capabilities, we evaluate several distinct proactive dialogue strategies on different components of trust. For a general understanding of the relation between proactive strategies and trust, we set up a mixed-factorial study which examines how different levels of proactivity as well as timing strategies of intelligent decision guide are trusted by subjects in an artificial game scenario. In order to provide insight into the how proactivity was modelled for our approach, we describe a formalisation of proactive dialogue in the following section.

III. FORMALISATION OF PROACTIVE DIALOGUE
According to McTear [61] and Litman and Pan [62] conversational interactions with current systems can be divided into three types in terms of the dialogue initiative strategy. In *user-directed dialogues* the user initiates and controls the dialogue. However, this strategy mostly supports

FIGURE 2. The IP-continuum ranging from zero to full autonomy [33].

so-called one-shot interactions, where the user issues a question or command and the system reacts. Hence, this strategy is typically applied in smart speakers and smart home assistants. In contrast, *system-directed* dialogues are led by the system. Here, McTear distinguishes between three types: the system initiates an interaction to deliver a reminder or notification; instructional dialogues in which the user starts the dialogue and the system repeatedly guides the user with little input from the user; so-called slot filling dialogues, in which a user commences the dialogue with requesting a service and the system takes over the command of the interaction posing a set of questions for working out the user's preferences and helping with task completion. The third strategy is known as *mixed-initiative dialogue*, that may not be confused with the term mixed-initiative interaction for human-machine collaboration on a specific task. Here, mixed-initiative refers to a system's ability to ask open-ended and specific questions, while providing the user with more freedom when answering questions in tasks-oriented dialogues [62]. Otherwise, the term can also be applied in open-domain interactions where the conversation can span a variety of topics and both the system and the user have control over the dialogue flow [61].

In this article, we focus on a mixed-initiative interaction where a user cooperates with a DIY-assistant for the execution of home improvement tasks. The basic interactions with the assistant follow the type of instructional dialogues, i.e. the system provides step-by-step descriptions for successful task execution. Therefore, the overarching dialogue flow follows a *system-directed pattern*. Hence, the system sequentially presents a task step and the available interaction choices (represented as a system action s_i), whereas the user can take different actions (represented as a user action u_i) until the collaboration ends after n task steps:

$$(s_1, u_1), (s_2, u_2), (s_3, u_3), \ldots, (s_n, u_n) \qquad (1)$$

For investigating the relations of the different proactive strategies and HCT, a decision-making use case scenario is considered. Here, a DIY-assistant helps the user in the planning of a specific DIY-project, i.e. supports selecting appropriate actions on how to perform individual task steps. In comparison to a simple instructional dialogue represented in Eq. 1, an agent is able to provide assistance before a user executes an action u_i. The assistance is provided either in a proactive, i.e. initiated by the system, or in a reactive manner, i.e. initiated by the user.

In the context of mixed-initiative interaction, proactive behaviour implies that the CA suggests or takes over actions

on behalf of the user. Therefore, proactive actions *pa* can be considered as the initiation of sub-dialogues, where the assistant influences a user action *u*, i.e. the system would interact with the user in order to affect their decision-making. Subsequently, a proactive action can be defined as a function of *u*, noted as *pa(u)*. Under this consideration, the structure of a system-directed dialogue can be updated as follows:

$$(s_1, pa(u), u_1), \ldots, (s_n, pa(u), u_n) \qquad (2)$$

Our preliminary work [12] dealt with the development of a set of proactive dialogue actions (*None*, *Notification*, *Suggestion*, *Intervention*). In this article, the way in which a system can interact with the user in order to affect their decision-making is represented by this set of actions. Hence, the proactive actions *pa* can be substituted with these proactive dialogue actions.

The set of actions was designed following the principles of the IP-continuum developed by Isbell and Pierce [33]. As previously described, the continuum ranges from zero ("Do It Yourself") to full automation ("System Makes Decisions") or from no intervention (not obtrusive) to complete intervention (highly obtrusive). Transferring the continuum to application in human-computer dialogue we summarised the second and third point of the continuum (see Fig. 2) under the proactive action *Notification*. The proactive actions content is modelled according the guidelines provided by Yorke-Smith [32], e.g. only task specific information is conveyed that contributes to the user's interests and tasks, or the system is aware of the current situation during task execution and aware of the user's current and future needs. Furthermore, proactive explanations can be added to justify the behaviour of the system to take the initiative. This fulfills Yorke-Smith's guideline to enhance the proactive system's understandability. Besides, justification explanations have shown to improve the user's trust in automatic systems [63]. Based on these considerations, we obtain four levels of proactivity:

A. NONE

This strategy refers to reactive system behaviour and forms the lowest level of proactive behaviour. In this condition, users can only explicitly request help from the assistant.

B. NOTIFICATION

This strategy forms the most conservative proactive approach. Following such a policy, the participant is only notified by the system. In this case it was up to the user to get assistance or to ignore the system's offer. By applying a notification,

the user is in control of the system's proactivity and is able to ignore it. However, this proactive action might shift the user's focus to possible helpful resources and might be perceived as unobtrusive.

C. SUGGESTION

Using aforementioned strategy, the CA directly suggests a solution by also providing a proactive explanation for its decision. Hence, the system takes over some control of the interaction and asks the user to make a choice. This forms a more rigid way of user interruption, but still lets the user in control over the final decision. As response to the system's proposal, a subject can either confirm or decline the suggestion.

D. INTERVENTION

In this case the system takes over all responsibilities and performs a particular action in place of the user, also providing a proactive explanation. Utilising this strategy might be perceived quite obtrusive, but can be helpful if the user has reached a critical level of need for proactivity.

Depending on the use case, this proactive dialogue framework forms the basis for developing proactive dialogue strategies. In the following section, we describe the implementation of these proposed proactive actions into a proactive dialogue companion for task-planning. Here. a task-planning scenario in the DIY-context forms an adequate way to test proactive strategies in a vulnerable environment which requires the user's trust in the assistant.

IV. A PROACTIVE DIALOGUE COMPANION FOR TASK-PLANNING

A. USE CASE SCENARIO

For developing and evaluating proactive dialogue strategies, a use case in the DIY domain was chosen. Even though the proactive dialogue strategies are generally domain-independent, this domain was selected as our previous work dealt with the development of an intelligent personal assistant for assisting users in DIY-tasks [4], [64]–[67]. For being applicable in other domains, only the content of the proactive actions needs to be adjusted for fitting the respective use case. The user's task in the test scenario was to plan two separate DIY-projects: building of a wooden nesting box and the assembly of a wall candle holder made from copper tubes. The projects differed in the degree to which they were familiar to the users and might effect their perception of the difficulty of the task. While building a nesting box was ought to be more known to users, a copper-tube wall candle holder was supposed to require a higher degree of imagination from the subjects and could hence be perceived as more difficult. Each project consisted of a predefined set of five sub-tasks ($n = 5$). The building of the wooden nesting box comprised the steps "wood cutting", "pre-drill holes", "connect the parts of the nesting box", "create an entrance hole", and "process wood". In contrast, the steps for the wall candle

holder were "saw copper tubes", "connect copper tubes", "polishing copper tubes", "pre-drill wall and dowell", and "attach wall candle holder".

For each task, the user had to make decisions on how they would perform individual task steps, without actually working on the task. They only had to select between different pre-defined approaches or tools which could help solving the task step. These option were presented on a task screen. The order of the task steps was fixed and could not be changed by the user. For each step, four options on how to accomplish the task were presented.

FIGURE 3. Screenshot of the interface for the planning task [12]. Users could choose between four different methods for task completion. All options were presented textually and visually. The selection was made either by clicking on the respective button or by confirming NAO's proposal.

An example of possible options for the sub-task "connect the parts of the nesting box" is depicted in Fig. 3. Subjects were told to select the options they considered best and that an intelligent personal assistant is able to help with decision-making, if necessary. Additionally, an artificial rewarding model was implemented for better motivating participants to engage in the task and to provide a risky environment in which trust is important. Therefore, options were associated with a rewarding model based on three fictional categories: appropriateness to the task, effort, and time efficiency. Each category was rated between 0 and 10 scoring points. The most common approach to perform a task was awarded the highest scoring (30). Alternative approaches that were functional but more cumbersome or effort intensive were awarded 0, 10, or 20 depending on their usefulness. In the example depicted in Fig. 3 the cordless screwdriver was the best option, while the usage of the nail gun was rated as inappropriate tool for this task and rated with 0 points. After selection, the score of the chosen approach was presented to the user as direct feedback.

For providing a deeper level of assistance, the intelligent agent was augmented with the ability to express proactive behaviour. Hence, in the DIY-scenario the user was accompanied actively or in a reactive manner by a personal assistant which would assist with decision-making in different ways. The assistant was designed to be an expert avoiding the unintended side effects of incompetent system behaviour on its trustworthiness. Thus, it would only suggest the most suited options. This allowed to only consider the effects of the proactive levels on the HCT.

For our experimental setup, we determined to let the participant control the planning task via a laptop that could be operated by mouse clicks on a purpose-built GUI. As an external representation of our assistance system was deemed
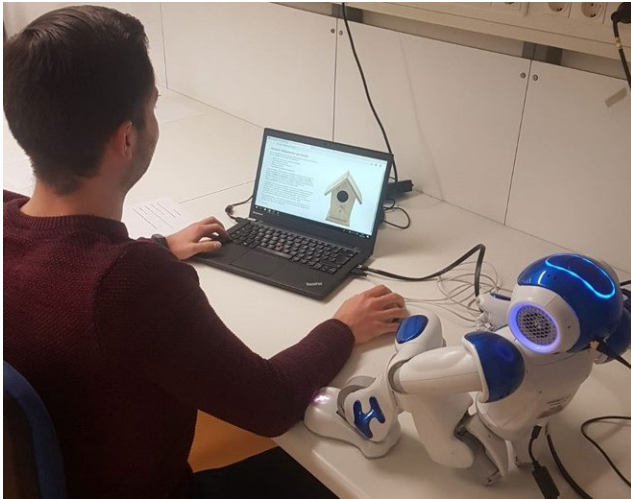
FIGURE 4. The robotic assistant NAO was placed in front of the task screen right beside the subject [12]. In doing so, the robot should be perceived as a team member for task completion. Additionally, NAO has been seated during the interaction for a better quality of speech recognition.

FIGURE 5. Depiction of the study apparatus.

to form a better separation of task and assisting technology, we have positioned a NAO robot next to the laptop. This 120 cm high humanoid robot from Softbank Robotics has an integrated speech interface which enables a natural approach to dialogue control. Hence, the setup seemed more realistic and was expected to deliver more significant results. To avoid confusing the user and to achieve a better quality of the speech recognition, the robot had a fixed position and deactivated autonomous movements. The experimental setup is illustrated in Fig. 4. The user was instructed which phrases can be used, e.g. "*Which option do you recommend in this situation?*". However, if a user input was not recognised, the system automatically requested a repeat.

B. SYSTEM DESIGN

For conducting the user study, we implemented a prototypical system consisting of a task-interface, a domain and reward model, and a DIY-assistant in form of the NAO robot. The study apparatus is visualised in Fig. 5. The user interface presenting content for the planning task was implemented as a clickable web application using the JavaScript framework with a Bootstrap plugin for designing the web pages. Fig. 3 shows a screenshot of the designed interface. The web page was structured in such a way that the description of the sub task was presented on top of the screen, whereas the four different options were put in line below the assignment. Each option was presented with a picture of the tool or approach and the corresponding label. NAO's proactive messages and responses to user requests were provided as spoken utterances using natural language. At each task step, the user could select one option using a mouse click and/or have a spoken dialogue with the robot for getting guidance on decision-making.

The domain model contained the content of the individual task steps, options as well as the content of the assistance
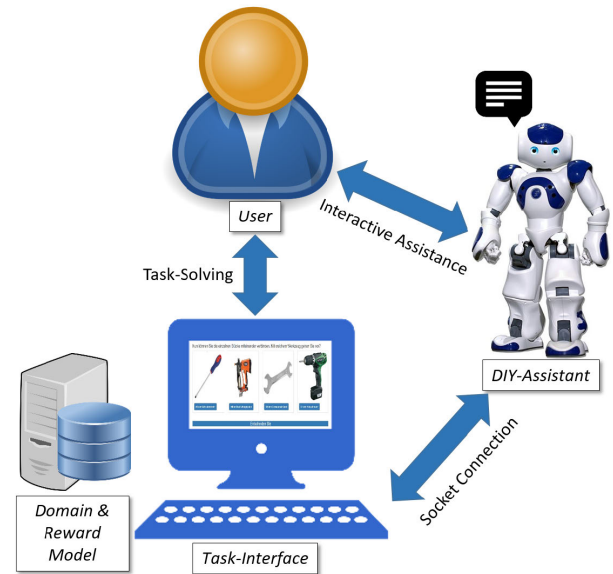
messages. The corresponding texts and images were pre-defined in advance and stored as html-templates. Similarly, the reward model was pre-defined while the association of the scoring points to the respective options was carried out relying on DIY-knowledge from internet research. The associations between scoring points and options were implemented using key-value pairs.

To obtain knowledge about when and how to provide the proactive messages, we have connected our external assistance system NAO to the web interface using NAO's QiMessaging developing framework. QiMessaging makes use of JavaScript bindings in order to take hold on NAO's speech modules. The bindings provide the class QiSession that connects to the robot via socket connections and gets proxies to services. After creating a session, NAO's modules (services) can be called using the service() function. This provides a JavaScript proxy to any service. These services are JavaScript object exposing methods and signals. A service method, e.g. the "AlTextToSpeech" method allowing NAO to provide speech output, is completely asynchronous. This enabled the interaction of NAO to be proactively initiated through timeouts on a web interface. Additionally, we implemented the "ALSpeechRecognition" service method for setting the language and vocabulary of the assistant's speech recognition. We provided a rich vocabulary for ensuring the recognition of multiple paraphrases of the statements the user was allowed to utter. For creating a system response upon the recognised user input, NAO's internal memory "ALMemory" was used. This memory provides callbacks on specific events, e.g. when speech was recognised, The event for speech recognition was subscribed to by our agent in order to react appropriately to speech commands. For example, in case the user uttered 'Which option do you recommend in this situation?' at previously described task step "connect the parts of the nesting box",
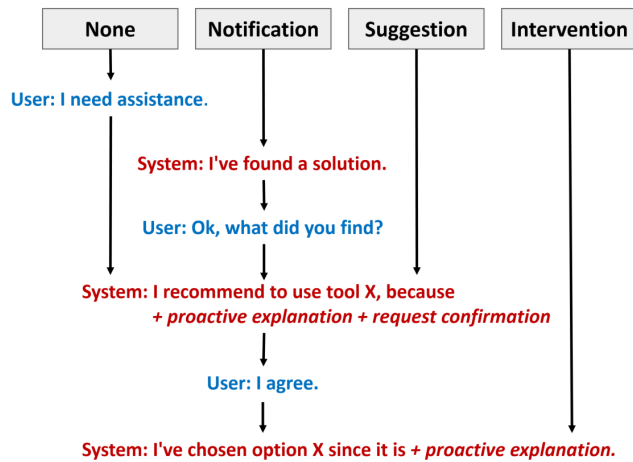
FIGURE 6. Flowchart, visualising the dialogue content of different levels of proactivity. User utterances are coloured in blue, while system actions are red-coloured [12].

the assistant would provide the suggestion "*The solution with the cordless screwdriver sounds good, because it is the most time efficient way. Should we choose this solution?*". As a response, the user could either accept or decline this offer using speech that would also trigger a speech event. How the timing and proactive actions of NAO were designed and implemented is described in detail in the following section.

C. DEVELOPMENT OF PROACTIVE STRATEGIES

Proactive assistance was modelled according to the proactive dialogue actions defined by [12]: *None*, *Notification*, *Suggestion*, *Intervention*. Since the scenario was a planning task and the user was supposed to select the, in their opinion, best option per individual task step, the purpose of system behavior was to provide helpful information and suggestions for the selection process via natural language. Therefore, we adapted the general proactive actions presented in the previous chapter to fit our use case. The explanations accompanying the proactive messages were generated using scripted templates.

Using the reactive *None*-action, the system waited for the user to explicitly ask for suggestions. For example, a user could say "*NAO, help me.*" for receiving assistance. As a response, the robot would then suggest the solution with the highest score, which is equivalent to the *Suggestion*-action. The more conservative proactive actions *Notification* and *Suggestion* let the user confirm the assistant's proposals and differ only in the degree of directness. While *Notification* allows the user to ignore the system's message and proceed on their own, the *Suggestion*-action expects the user to accept or decline the offer. When users reacted upon a system's notification, a suggestion was triggered. The *Intervention*-actions took the responsibility completely out of the user's hands an autonomously chose an option. Here, the system makes an selection for the user and would utter: "*I have chosen the solution with the cordless screwdriver, because it is the most time efficient*". In parallel, the option is chosen

and the user was lead to the next task step. Possible dialogue flows of the proactive strategies are depicted in Fig. 6. The reason why the *Suggestion*-strategy was used, either upon user request or after the user had reacted to an active system notification, was to induce a natural interaction behaviour. If NAO's proposal was rejected by the user, the system did not engage any further proactive interaction at this task step.

For triggering the proactive system's actions, we made use of timeouts. This allowed for specifying a certain period of time, after which the robot took the initiative. The rules upon we implemented the timing of system actions are explained as follows: A *fixed* timing strategy was used as a baseline. For this purpose, we hard-coded the points of time the system took the initiative during the execution of the planning task. Of the five possibilities for taking the initiative, the system proactively acted at the sub-tasks one, four, and five. In doing so, a "quasi"-random proactive system behaviour was simulated. Actually randomly distributing the timing was omitted for economical reasons. A randomisation would have enlarged the possibilities for triggering proactive behaviour, requiring a huge amount of subjects in order to guarantee comparability. Technically, NAO took the initiative eight seconds after the respective task screen was loaded. To avoid that subjects could select an option before NAO had behaved proactively, we blocked the selection buttons for this period of time. As a cover up, participants were told that we wanted to guarantee that they really have read and understood the task.

Furthermore, we implemented an *insecurity*-based timing strategy. Here, the robot could take the initiative at each project step, if the subject had not requested help or had not selected an option before a time limit of twelve seconds. We interpreted the four seconds of user inactivity after the selection buttons had been enabled as insecurity in task performance. As indication of insecurity is extremely user-dependent, this time period was chosen as a heuristic measure based on pre-testing. By using the participant's insecurity during task performance, a more context-related strategy for timing of proactive behaviour was provided.

D. HYPOTHESES

We expected the developed proactive dialogue strategies to perform differently with respect to the HCT relationship. In particular, low- to medium-level of proactivity (none, notification, and suggestion) were supposed to lead to higher ratings of trust than high-level proactivity (intervention) due to the uncontrollableness of the autonomous system's actions. Additionally, we expected a medium-level of proactivity (notification and suggestion) to perform better regarding trust than low- and high-level proactivity, as a result of the benefits of proactive actions outweighing the costs. Furthermore, we expected proactive dialogue actions to lead to higher ratings of trust than the reactive *None*-action for the timing strategy based on the user's insecurity in contrast to the fixed timing strategy. Besides, medium-level of proactivity
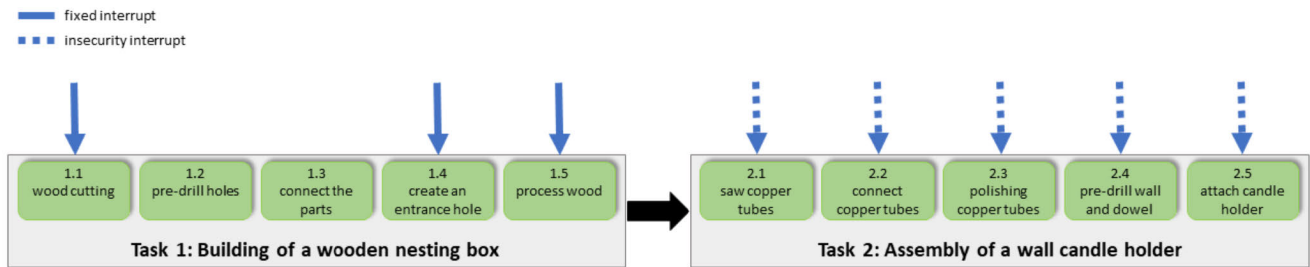
FIGURE 7. Schematic description of the study procedure [12]. Both planning tasks consisted of five sub-tasks. For the strategy "fixed interrupt", the system intervened proactively in the sub-tasks 1, 4, and 5. For the strategy "insecurity interrupt", the system intervened proactively each time a user insecurity was detected. The two strategies were switched depending on the test condition.

was ought perform better regarding trust than high-level proactivity based on the user's insecurity.

V. EXPERIMENTAL DESIGN

A. CONDITIONS
In our study setup a 2×4 mixed factorial experimental design was conducted with proactive dialogue strategies (none - notification - suggestion - intervention) as between-independent variables. Moreover, the timing strategies for proactive behaviour (fixed - insecurity-based) were used as within-subject variables. The order of the timing strategies was randomised for each proactive dialogue strategy except for the none condition, which did not require any timing due to reactive behaviour. Accordingly, this study setup resulted in a randomised distribution of participants in seven study groups. The order of the tasks was the same for all users.

B. PARTICIPANTS
42 German participants (50 % female) with an average age of 26 (standard deviation (SD) = 4.15) were recruited and received 10 Euro as a reward. Most subjects were students (37) majoring either in psychology (27 %) or in computer science (38 %).

C. EXPERIMENTAL PROCEDURE
After the welcome procedure, participants were provided with first instructions and details about the study. As a cover story they were told that the purpose of the study was to test a decision-making algorithm of the NAO robot and to generally consider problem-solving between humans and robots. Afterwards, they had to read and sign the informed consent and had to fill out a pre-test questionnaire. Before the first interaction cycle, they received detailed information about the tasks and the procedure of the study. This included details about the speech capabilities of NAO and about the task to rate the interaction with the robot. Subsequently, the participants had to work on planning the first DIY-project. After completion, they had to fill in a questionnaire to assess the dependent variables and to check the manipulations. The same procedure was repeated for the second task scenario. In addition, the questionnaire of the second task also contained an evaluation of overall perceived user experience with the robotic system. In conclusion, participants received

their reward and were dismissed. A graphical representation of the procedure is depicted in Fig. 7.

D. QUESTIONNAIRES
In our experiment, we assessed trust and its five bases (competence, reliability, understandability, personal attachment, faith) in the robot and the participants' cognitive load during the interaction in order to evaluate the effects of proactive dialogue behaviour. Furthermore, we measured the user's experience with the system in general for checking the quality of the setup. Each variable was measured with items from established and validated scales. To determine trust towards the robot, a short version of the Trust in Automated Systems Scale [59] in German by Kraus [60] was implemented. The scale consists of 7-items for measuring the user's trust and showed great performance for investigating learned trust during and after interaction. Furthermore, scales for measuring the bases of trust developed by Madsen and Gregor [39] were used. For measuring three types of cognitive loads (extraneous, germane, intrinsic), a questionnaire developed by Klepsch *et al.* [68] was included. More information on the different types of cognitive load is described in [69], for example. The user's experience with the system was assessed via the user experience questionnaire (UEQ) developed by Laugwitz *et al.* [70]. This questionnaire consists of six word-pair scales measuring the pragmatic (Perspicuity, Efficiency, Dependability) and hedonic qualities (Stimulation, Novelty) of a technical system. Besides, for personality assessment, the Big-Five-Inventory BFI-10 by Rammstedt *et al.* [71] was included. The scales, which were only available in English language, were translated into German. Besides, some scales were slightly modified for content and study context. For example, we clarified that participants rate the interaction with the NAO-robot and not the task-interface, as the original questionnaires make use of the neutral term "system" in the scale statements.

Possible confounding variables were measured using scales of predisposed trust in autonomous systems [72], negative attitudes towards robots (NADAS) [73], as well as self-developed scales for previous experience with speech dialog systems and DIY-tasks. All scales were rated on a 7-point Likert-scale from 1 (strongly disagree; word adjective (UEQ)) to 7 (strongly agree; word adjective (UEQ)).

TABLE 2. Descriptive statistics of the UEQ-subscales and the overall perceived trust in the system with reference to the proactive dialogue actions. Mean values (M) and standard deviations (SD) are presented.

Proactive Action		Attractiveness	Perspicuity	Novelty	Stimulation	Dependability	Efficiency	Trust
		M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
None	overall	5.23 (.98)	5.65 (1.32)	4.73 (.89)	4.83 (.53)	5.73 (.79)	5.08 (1.14)	5.52 (.87)
	female	5.22 (1.12)	5.50 (1.35)	4.75 (.59)	4.96 (.53)	5.75 (.57)	5.41 (.92)	5.70 (.78)
	male	5.25 (.87)	5.88 (1.44)	4.69 (1.33)	4.63 (.52)	5.59 (1.14)	4.56 (1.39)	5.25 (1.05)
Notification	overall	4.97 (.89)	6.18 (.73)	5.32 (.62)	5.11 (.81)	5.73 (.66)	5.55 (1.02)	6.25 (.61)
	female	5.21 (1.16)	6.56 (.52)	5.50 (.74)	5.63 (.88)	6.00 (.74)	6.13 (1.09)	6.52 (.45)
	male	4.83 (.78)	5.96 (.77)	5.21 (.59)	4.82 (.66)	5.57 (.61)	5.21 (.88)	6.09 (.67)
Suggestion	overall	4.76 (.60)	5.00 (.93)	4.68 (.81)	4.57 (.73)	5.14 (.85)	5.05 (.94)	5.82 (1.14)
	female	4.86 (.39)	5.46 (.82)	4.68 (.98)	5.00 (.48)	5.04 (.78)	5.39 (.93)	5.72 (1.05)
	male	4.58 (.92)	4.19 (.38)	4.69 (.55)	3.81 (.38)	5.31 (1.05)	4.44 (.66)	5.98 (1.43)
Intervention	overall	5.03 (.88)	4.93 (1.35)	4.33 (1.14)	4.68 (1.09)	5.08 (1.06)	5.18 (.92)	5.64 (.94)
	female	5.08 (1.06)	4.69 (1.61)	3.50 (1.29)	4.38 (1.01)	4.94 (.75)	5.31 (.83)	6.27 (.34)
	male	5.00 (.85)	5.08 (1.29)	4.88 (.65)	4.88 (1.19)	5.17 (1.29)	5.08 (1.04)	5.21 (1.00)

VI. RESULTS

For data analysis, we used t-tests for the manipulation checks, a multivariate analysis of variance (ANOVA) for confounding variables, as well as a mixed ANOVA for testing the significance of developed proactive dialogue strategies. No significant outliers were found in the data set. Typically for small sample sizes, some deviations of normal distribution (Shapiro-Wilk test) were detected. However, the F-statistic of an ANOVA can be robust to violations of assumptions, as long as the group sizes are equal [74]. Furthermore, we tested for homogeneity of the error variances using Levene's test. All dependent variables showed equal variances ($p > .05$) except the measurement of understandability which was then excluded from the ANOVA.

A. CONFOUNDING VARIABLES AND MANIPULATION CHECK

Confounding group differences for proactive behaviour could be ruled out as the multivariate ANOVA did not reveal any significant differences (all p-values ≫ .05). The evaluation of the manipulation check confirmed the successful manipulation of proactive dialogue behaviour (all p-values < .05 with regard to the non-proactive strategy). However, the manipulation of timing of proactive behaviour dependent on the subject's insecurity failed (all p-values ≫ .05). Therefore, we did not consider timing strategies in the evaluation. Instead, we checked if the two tasks differed significantly in their level of difficulty. The conduction of a paired t-test revealed that the intrinsic cognitive load, related to the difficulty of a task, was rated significantly higher for the planning of the wall candle than for the planning of the nesting box (mean (M) = 1.94, SD = 1.08 for nesting box vs. M = 2.48, SD = 1.20 for wall candle, $t(41) = -3.46, p < .01$). Hence, differences between proactive dialogue actions depending on task difficulty could be observed.

B. USER EXPERIENCE WITH THE EXPERIMENTAL SETUP

In order to ensure the functionality and usability of employing NAO as an intelligent assistant, we evaluated the system regarding user experience. In general, the system received positive feedback. Participants rated their interaction partner well understandable, represented by a high value for perspicuity (M = 5.45, SD = 1.18). Furthermore, the system received good ratings for dependability (M = 5.42, SD = .87) and efficiency (M = 5.21, SD = .99). In addition, the interaction with NAO received moderately good ratings for attractiveness (M = 4.99, SD = .83), novelty (M = 4.77, SD = .92), and stimulation (M = 4.80, SD = .87). The descriptive statistics of the UEQ-subscales with reference to the proactive dialogue actions are depicted in Table 2. The *Notification*-action was rated higher than the *Intervention*-action for the UEQ-features perspicuity ($t(13.48)$ = 2.61, p < .05) and novelty ($t(19) = 2.52, p < .05$). Furthermore, the *Notification*-action was rated higher than the *Suggestion*-action for perspicuity ($t(20) = 3.33, p < .01$). The comparison of the proactive action showed no significant differences on any other UEQ-subscale.

C. EFFECTS OF PROACTIVE DIALOGUE STRATEGIES ON TRUST

There was a statistically significant interaction between proactive dialogue actions and task difficulty for perceived competence ($F(3, 38) = 8.25, p < .001, \eta^2 = .39$) and for perceived reliability ($F(3, 38) = 3.95, p = .015, \eta^2 = .24$). In order to investigate further which groups differed significantly in which task, a series of t-tests with Bonferroni correction was conducted. First, we examined the effects of proactive actions on perceived competence. The *Notification*-action was evaluated significantly higher than the *None*-, and *Intervention*-action for the task *nesting box* ($t(19) = 4.46, p < .001$ vs. None; $t(19) = 2.93, p = .038$ vs. Intervention). Furthermore, for the more difficult task *wall candle* the *Notification*-action was rated higher than the *Intervention*-strategy ($t(19) = 2.90, p = .038$). In the following, results for perceived reliability are presented. For the relatively easier task *nesting box*, the *Notification*-action was graded significantly higher than the *None* ($t(19) = 3.03, p = .028$ vs. None). For the harder task *wall candle* the
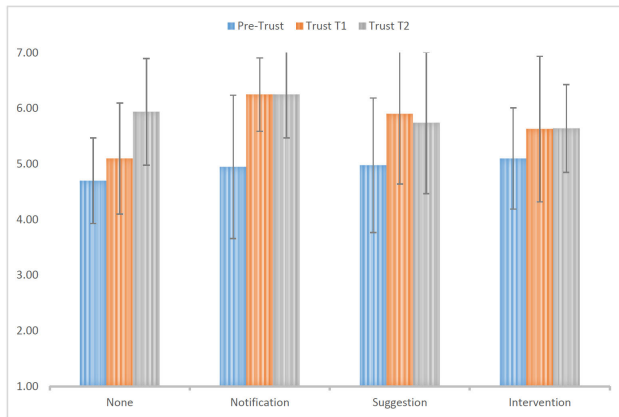
FIGURE 8. Trust progression over the course of the experiment with regard to the proactive dialogue actions. Pre-Trust represents predisposition to trust autonomous systems, while Trust T1 and T2 represent the trust measurements after the tasks "nesting box" and "wall candle holder" respectively.

Notification- and *None*-action were rated significantly higher than the *Intervention*-action (Notification vs. Intervention, $t(19) = 2.96, p = .032$; None vs. Intervention, $t(18) = 2.84, p = .044$). These results are depicted in Fig. 9. Finally, we investigated significant main effects of proactive dialogue strategies. The *Notification*-action was evaluated higher than the *Intervention*-action for the categories perceived competence ($t(19) = 3.02, p = .028$) and perceived reliability ($t(19) = 3.16, p = .020$). Additionally, the *Notification*-action was rated higher than the *None*-action for perceived competence ($t(19) = 2.00, p = .036$). No significant results were found for all of the remaining dependent variables.

Considering the trust progression over the course of the experiment depending on the proactive actions, we investigated the within-subject differences of the trust ratings before the experiment and after each task. Hereby, initial trust was measured using the predisposed trust in autonomous systems scale. For testing the significance of the differences, we used paired t-tests. We found a significant trust difference for the *None*-action measured after the first and second task ($t(9) = -3.00, p = .015$). Furthermore, we found significant trust differences between initial trust and trust measured after the first task for the actions *Notification* ($t(10) = -3.95$, $p = .003$) and *Suggestion* ($t(10) = -2.90, p = .016$). There was no significant trust progression for the *Intervention*-action. The results are depicted in Fig. 8.

D. EFFECTS OF USER CHARACTERISTICS ON TRUST
Further exploring the data, we found significant gender differences using t-tests on the independent samples. Females rated themselves to be less experienced with DIY ($t(40) = 2.13, p = .039$). Additionally, they rated themselves notably less experienced with CAs ($t(40) = 1.94, p = .059$). Considering personality characteristics, females had higher ratings for neuroticism ($t(40) = -3.33, p = .002$) and conscientiousness ($t(40) = -2.22, p = .032$). Females

rated themselves also considerably more open to experiences ($t(40) = -1.83, p = .075$).

For observing the effects of the proactive actions depending on the individual gender, we split the data set accordingly and tested for significant differences. Due to the resulting smaller sample size, a normal distribution of the data was not further provided. Therefore, we utilised a Kruskal–Wallis one-way analysis of variance for testing the effects of the different proactive actions. Here, several significant differences were found for the female gender. For the task "nesting box", significant differences were found for reliability ($p = .042$) and competence ($p = .038$). A significant difference was found for the task "wall candle" regarding reliability ($p = .005$). Additionally, we found a significant effect on overall perceived reliability ($p = .020$).

In order to investigate further which groups differed significantly for each task, post-hoc tests using the Dunn-Bonferroni method were conducted. The results showed that the *Notification*-action was rated higher than the *None*-action for reliability ($Z = -2.83, p = .028$ and competence ($Z = -2.88, p = .024$ in the task "nesting box". In the task "wall candle", the *None*- and *Notification*-action were rated higher than the *Intervention*-action for reliability ($Z = -3.18, p = .009$; $Z = -2.95, p = 0.019$). Additionally, we found some notable effects of the proactive actions on competence ($p = .055$) and understandability ($p = .059$) for the task "wall candle" and for the UEQ-dimensions novelty ($p = .090$) and dependability ($p = .093$). For the male gender no significant differences were found. However, a notable effect of the proactive actions on the UEQ-dimension perspicuity was found (($p = .070$).

VII. DISCUSSION
The study results verified our hypotheses that altering the degree of proactive system behaviour has a significant impact on the user's trust in the CA. Especially, we discovered interesting insights into the relations between proactive actions and task knowledge/difficulty as well as user characteristics on the perceived competence and reliability of the system. For the first, easier perceived task "nesting box", low- and medium-level proactive system actions were particularly trusted more than the reactive condition. This was validated both by the examination of the trust progression analysis and the ANOVA. For this, there exist two possible explanations.

First, users could have been more sure about the decision on appropriate planning steps for this task in comparison to the "wall candle task". Therefore, the proactive actions could have been perceived as a confirmation or reinforcement of their own decision making processes and relieved them in task execution. This in turn could foster trust, as the benefits of proactive actions were higher as compared of the risks of wrongful system advice. Particularly, as the low- and medium level are more controllable [33]. The relative low ratings for competence and reliability of the *None*-action for the "nesting box"-task could be explained that the ratio
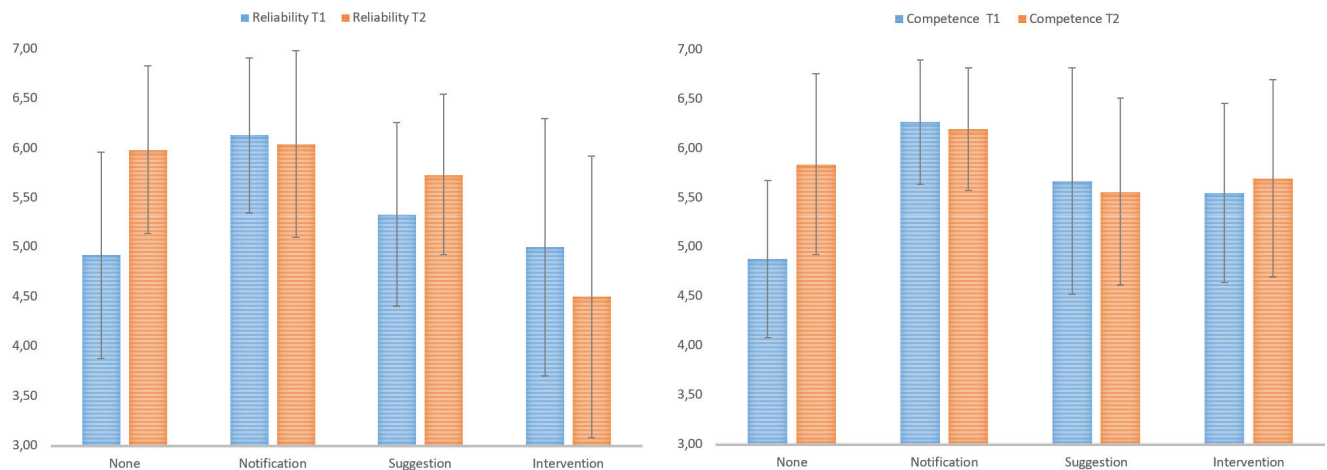
FIGURE 9. Depiction of the results for perceived reliability (left) and perceived competence (right) depending on the four proactive dialogue strategies (Intervention, none, notification, suggestion) and the two tasks (1 = 'nesting box'; 2 = 'wall candle'). Mean values and standard deviations are provided.

between expenses and benefits of system usage was too low, as requesting the system for help was perceived as unnecessary step and could have been more of a distraction. For the second, more difficult task, the *None*-action was trusted similarly to the medium-level proactive conditions, as the benefits of requesting system help outweighed the costs of addressing the system.

Another explanation could be that study participants perceived low- and medium level proactive actions to help better in getting familiar with the task and the CA's design and performance than a reactive system that does not actively communicate. Hence, the dynamically learned trust according to Hoff and Bashir [50] was increased more by proactive actions in the first task, because they initially made the system more transparent. For the second task, the *None*-action increased the dynamically learned trust as subjects started communicating more with the system and learned about its benefits.

Among the proactive strategies, the *Notification*-action had the most impact on conveying competence and reliability of the system. This particularly holds true for assistance in the first, easier task. The *Notification*-strategy formed the most conservative proactive strategy, that offered help in a more subtle way. Hence, study participants always felt in control but were also aware of the system's active assistance. Furthermore, this strategy comprised the most (four) dialogue turns. It seems that subjects tend to accept proactive system behaviour more when it is possible for them to have natural dialogues.

In line with the findings from previous work by Rau *et al.* [11], the most autonomous system behaviour, the *Intervention*-action, was less trusted than the more conservative strategies. Subjects considered this strategy to be too obtrusive and perceived the system to be imposing. In summary, when being proactive, a system should act more subtle and give the user a feeling of system involvement in the task, i.e. by notifying about or suggesting information.

The *Intervention*-strategy could be used for really tedious or annoying tasks. Therefore, we consider the *Notification*- and *Suggestion*-strategy as more trustworthy for the user. These findings reinforce the results by Peng *et al.* [2], designating medium-level proactivity as the most helpful.

When considering how user characteristics affect the relation of proactive system actions and HCT, we found significant gender differences. The interplay between gender and trust is a common phenomenon in engineering and science [75]–[77]. We found that varying the degree of proactive system behaviour had a particularly significant impact on the females user's cognitive-based trust, reliability and competence. Female study participants were less experienced with CAs and DIY than male subjects. This suggests a first evidence, that the perception of proactive system behaviour as trustworthy is crucially affected by the user's experience with the task and technology. Furthermore, we found significant differences between the genders regarding the big five personality traits as females rated themselves higher for neuroticism, conscientiousness, and to some degree openness towards new experiences. A high degree of openness to experience relates curious, innovative, adventurous persons. A high degree of conscientiousness relates to goal-oriented, efficient, disciplined, organised behaviour. Sensitive, insecure individuals can be related to have a high degree of neuroticism. Examining the individual personality traits it could be reasoned that proactive behaviour primarily affects innovative, goal-driven, but also more insecure persons. Interestingly, in organisational psychology and management, proactive behaviour is associated with goal-directed activities and innovation [14], [15] relating to the traits openness and conscientiousness. Seibert *et al.* [78] also introduce the "proactive personality". Hence, there could be a correlation between one's tendency for proactive behaviour and the perception of a proactive CA. However, more research on this topic is necessary for providing clear insights and underpin this hypothesis. Nonetheless, taking into consideration the

user's personality when developing a proactive CA could be beneficial.

The reason why overall trust in the system did not differ significantly could lie in the short duration of the interaction. These kinds of interactions only influence the cognitive-, and not the affect-based trust. In order to get significant differences in overall trust, a more long-term human-machine relationship is necessary. According to Madsen and Gregor [39], both cognitive- and affect-based trust must be perceived as high in order to establish an overall trustworthy CA.

However, manipulation of timing strategy according to the user's insecurity failed. In consequence, we assume that a time-dependent measure for insecurity is insufficient for usage as a trigger-variable of proactive dialogue actions. Arguably, time as initiation-criterion needs to be avoided at all because there exist too many side factors, which are not necessarily user-related, that could lead to a delay in time.

Nevertheless, our work had several limitations. Even though we let subjects interact with an actual autonomous system, the study was still conducted in a controlled environment. In a realistic scenario, a DIY-planning task would be much more unpredictable and unbounded. Additionally, NAO only allows for a limited speech interaction due to its technical constraints. Furthermore, the timing strategies developed can only be controlled in an experimental setup and can hardly be transferred to a real case scenario. Using the user's insecurity as a metric for timing proactivity, however, has proven unreliable. For tracking the user's insecurity, more sophisticated approaches, e.g. observing the mouse movement of a user or applying eye tracking, could present a more effective method. Since we kept using the same level of proactivity for a subject while going through a study run, this resulted in the perception of a rigid system. Even though the proactive dialogue strategies are in general domain and system independent making our results generalisable, their applicability needs to be tested in different domains in future work. Also a more diverse user group, i.e. a wider range of the subject's age. profession, or different cultures, could help with generalisation of the approach. Additionally, the proactive actions need to be tested in more flexible and dynamic task structures, in order to investigate if proactive behaviour can improve task efficiency.

VIII. CONCLUSION AND FUTURE WORK

Due to the increasing abilities of autonomous systems, proactive conversations will become a major design guideline for developing high-quality assistance behaviour. By equipping a system with proactive functionality, a system may be able to relieve a user in task execution and competently provide reliable guidance and advice. The key to conveniently convey proactive system actions is to engage in a dialogue. Therefore, this article provided insights in developing proactive dialogue strategies for CA. We described four proactive actions or levels for engaging in a dialogue for assisting users. The results of an initial study on the user

perception of these actions showed the overall benefits of low- to medium-level proactivity and its relations to the HCT relationship. Furthermore, we discovered an interaction of proactive actions and perceived task difficulty, as well as dependencies between proactive dialogue and certain user characteristics, such as domain experience, technical affinity, and personality properties. The results further showed that the low- to medium-level proactivity is better for establishing an immediate trust relationship. Particularly, proactive dialogue seems to be especially relevant for novice users. In future work, a more flexible way of proactive behaviour is intended, as we only considered one type of proactive action at a time in this work. For truly intelligent proactive assistance a dialogue system should be able to choose the appropriate type of action at the right time following an efficient proactive dialogue strategy. Therefore, we need to consider temporal relationship between the individual actions during different task steps. Here, we plan to implement a system that is able to learn an optimal proactive strategy. The systems should derive this strategy from real users, e.g. via reinforcement learning methods considering the user's trust state during the interaction. In doing so, we aim to create an even more trustworthy human-machine interaction.

REFERENCES

[1] Statista. (Aug. 2016). *Anzahl der Nutzer Virtueller Digitaler ASsistenten Weltweit in den Jahren von 2015 Bis 2021*. [Online]. Available: https://de.statista.com/statistik/daten/studie/620321/umfrage/nutzung-v%on-virtuellen-digitalen-assistenten-weltweit/

[2] Z. Peng, Y. Kwon, J. Lu, Z. Wu, and X. Ma, ''Design and evaluation of service robot's proactivity in decision-making support process,'' in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, p. 98.

[3] J. Paladines and J. Ramirez, ''A systematic literature review of intelligent tutoring systems with dialogue in natural language,'' *IEEE Access*, vol. 8, pp. 164246–164267, 2020.

[4] G. Behnke, P. Bercher, M. Kraus, M. Schiller, K. Mickeleit, T. Häge, M. Dorna, M. Dambier, D. Manstetten, W. Minker, and B. Glimm, ''New developments for Robert—Assisting novice users even better in DIY projects,'' in *Proc. Int. Conf. Automated Planning Scheduling*, vol. 30, 2020, pp. 343–347.

[5] K. Yoshino and T. Kawahara, ''News navigation system based on proactive dialogue strategy,'' in *Natural Language Dialog Systems and Intelligent Assistants*. Cham, Switzerland: Springer, 2015, pp. 15–25.

[6] J. Baraglia, M. Cakmak, Y. Nagai, R. Rao, and M. Asada, ''Initiative in robot assistance during collaborative task execution,'' in *Proc. 11th ACM/IEEE Int. Conf. Hum. Robot Interact. (HRI)*, Mar. 2016, pp. 67–74.

[7] F. Nothdurft, S. Ultes, and W. Minker, ''Finding appropriate interaction strategies for proactive dialogue systems—An open quest,'' in *Proc. 2nd Eur. 5th Nordic Symp. Multimodal Commun.*, no. 110, Tartu, Estonia: Linköping Univ. Electronic Press, Aug. 2015, pp. 73–80.

[8] J. L. Jenkins, B. B. Anderson, A. Vance, C. B. Kirwan, and D. Eargle, ''More harm than good? How messages that interrupt can make us vulnerable,'' *Inf. Syst. Res.*, vol. 27, no. 4, pp. 880–896, Dec. 2016.

[9] D. C. McFarlane and K. A. Latorella, "The scope and importance of human interruption in human-computer interaction design," *Hum. Comput. Interact.*, vol. 17, no. 1, pp. 1–61, Mar. 2002.

[10] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Trans. Comput.-Hum. Interact.*, vol. 12, no. 2, pp. 293–327, Jun. 2005.

[11] P.-L. P. Rau, Y. Li, and J. Liu, "Effects of a social robot's autonomy and group orientation on human decision-making," *Adv. Hum. Comput. Interact.*, vol. 2013, p. 11, 2013.

[12] M. Kraus, N. Wagner, and W. Minker, "Effects of proactive dialogue strategies on human-computer trust," in *Proc. 28th ACM Conf. User Modeling, Adaptation Personalization*, Jul. 2020, pp. 107–116, doi: 10.1145/3340631.3394840.

[13] P. R. Cohen, "Back to the future for dialogue research," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 9, Apr. 2020, pp. 13514–13519. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/7073, doi: 10.1609/aaai.v34i09.7073.

[14] J. M. Crant, "Proactive behavior in organizations," *J. Manage.*, vol. 26, no. 3, pp. 435–462, Jun. 2000.

[15] M. Frese and D. Fay, "4. Personal initiative: An active performance concept for work in the 21st century," *Res. Organizational Behav.*, vol. 23, pp. 133–187, Jan. 2001.

[16] S. K. Parker, H. M. Williams, and N. Turner, "Modeling the antecedents of proactive behavior at work," *J. Appl. Psychol.*, vol. 91, no. 3, p. 636, 2006.

[17] E. Horvitz, "Principles of mixed-initiative user interfaces," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. CHI limit (CHI)*, 1999, pp. 159–166.

[18] V. Balaraman and B. Magnini, "Proactive systems and influenceable users: Simulating proactivity in task-oriented dialogues," in *Proc. 24th Workshop Semantics Pragmatics Dialogue-Full Papers, Virually at Brandeis, Waltham, New Jersey, July (SEMDIAL)*, Jul. 2020, pp. 1–12.

[19] M. Kraus, F. Fischbach, P. Jansen, and W. Minker, "A comparison of explicit and implicit proactive dialogue strategies for conversational recommendation," in *Proc. 12th Lang. Resour. Eval. Conf.*, 2020, pp. 429–435.

[20] Y. Kato, T. Kanda, and H. Ishiguro, "May i help you?: Design of human-like polite approaching behavior," in *Proc. 10th Annu. ACM/IEEE Int. Conf. Hum. Robot Interact.*, Mar. 2015, pp. 35–42.

[21] A. Garrell, M. Villamizar, F. Moreno-Noguer, and A. Sanfeliu, "Teaching robot's proactive behavior using human assistance," *Int. J. Social Robot.*, vol. 9, no. 2, pp. 231–249, Apr. 2017.

[22] N. Wagner, M. Kraus, N. Rach, and W. Minker, "How to address humans: System barge-in in multi-user HRI," in *Proc. 10th Int. Workshop Spoken Dialogue Syst. Increasing Naturalness Flexibility Spoken Dialogue Interact.* Singapore: Springer, 2021, pp. 147–152.

[23] M. Awais and D. Henrich, "Proactive premature intention estimation for intuitive human-robot collaboration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 4098–4103.

[24] W. Y. Kwon and I. H. Suh, "Planning of proactive behaviors for human–robot cooperative tasks under uncertainty," *Knowl.-Based Syst.*, vol. 72, pp. 81–95, Dec. 2014.

[25] J. Grosinger, F. Pecora, and A. Saffiotti, "Making robots proactive through equilibrium maintenance," in *Proc. IJCAI*, 2016, pp. 3375–3381.

[26] P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro, "Learning proactive behavior for interactive social robots," *Auto. Robots*, vol. 42, no. 5, pp. 1067–1085, Jun. 2018.

[27] H. S. M. Cramer, N. A. Kemper, A. Amin, and V. Evers, "The effects of robot touch and proactive behaviour on perceptions of human-robot interactions," in *Proc. 4th ACM/IEEE Int. Conf. Hum. Robot Interact. (HRI)*, Mar. 2009, pp. 275–276.

[28] J. M. Beer, A. D. Fisk, and W. A. Rogers, "Toward a framework for levels of robot autonomy in human-robot interaction," *J. Hum. Robot Interact.*, vol. 3, no. 2, pp. 74–99, 2014.

[29] T. B. Sheridan and W. L. Verplank, "Human and computer control of undersea teleoperators," Massachusetts Inst. Technol., Cambridge, MA, USA, Man-Mach. Syst. Lab Rep., 1978.

[30] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 30, no. 3, pp. 286–297, May 2000.

[31] R. Sarikaya, "The technology behind personal digital assistants: An overview of the system architecture and key components," *IEEE Signal Process. Mag.*, vol. 34, no. 1, pp. 67–81, Jan. 2017.

[32] N. Yorke-Smith, S. Saadati, K. L. Myers, and D. N. Morley, "The design of a proactive personal agent for task management," *Int. J. Artif. Intell. Tools*, vol. 21, no. 1, Feb. 2012, Art. no. 1250004.

[33] C. L. Isbell and J. S. Pierce, "An IP continuum for adaptive interface design," in *Proc. HCI Int.*, Jul. 2005, pp. 1–10.

[34] S. T. Iqbal and B. P. Bailey, "Effects of intelligent notification management on users and their tasks," in *Proc. 26th Annu. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2008, pp. 93–102.

[35] W. Wu, Z. Guo, X. Zhou, H. Wu, X. Zhang, R. Lian, and H. Wang, "Proactive human-machine conversation with explicit conversation goals," 2019, *arXiv:1906.05572*. [Online]. Available: http://arxiv.org/abs/1906.05572

[36] K. Christakopoulou, F. Radlinski, and K. Hofmann, "Towards conversational recommender systems," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 815–824.

[37] H. Lopez-Tovar, A. Charalambous, and J. Dowell, "Managing smartphone interruptions through adaptive modes and modulation of notifications," in *Proc. 20th Int. Conf. Intell. User Interface*, Mar. 2015, pp. 296–299.

[38] J. F. de Abreu, R. Santos, T. Silva, T. Marques, and B. Cardoso, "Proactivity: The next step in voice assistants for the TV ecosystem," in *Proc. Iberoamerican Conf. Appl. Usability Interact. TV.* Cham, Switzerland: Springer, 2019, pp. 103–116.

[39] M. Madsen and S. Gregor, "Measuring human-computer trust," in *Proc. 11th Australas. Conf. Inf. Syst.*, vol. 53. Brisbane, QLD, Australia: Citeseer, 2000, pp. 6–8.

[40] M. Deutsch, "The effect of motivational orientation upon trust and suspicion," *Hum. Relations*, vol. 13, no. 2, pp. 123–139, May 1960.

[41] J. B. Rotter, "Interpersonal trust, trustworthiness, and gullibility," *Amer. Psychologist*, vol. 35, no. 1, pp. 1–7, 1980.

[42] J. K. Rempel, J. G. Holmes, and M. P. Zanna, "Trust in close relationships," *J. Personality Social Psychol.*, vol. 49, no. 1, p. 95, 1985.

[43] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Acad. Manage. Rev.*, vol. 20, no. 3, pp. 709–734, 1995.

[44] J. A. Simpson, "Psychological foundations of trust," *Current Directions Psychol. Sci.*, vol. 16, no. 5, pp. 264–268, Oct. 2007.

[45] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 39, no. 2, pp. 230–253, 1997.

[46] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[47] K. E. Schaefer, J. Y. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 58, no. 3, pp. 377–400, Mar. 2016.

[48] B. M. Muir, "Trust between humans and machines, and the design of decision aids," *Int. J. Man. Mach. Stud.*, vol. 27, pp. 527–539, Nov. 1987.

[49] B. M. Muir and N. Moray, "Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation," *Ergonomics*, vol. 39, no. 3, pp. 429–460, Mar. 1996.

[50] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Hum. Factors*, vol. 57, no. 3, pp. 407–434, 2015.

[51] D. Gorecky, M. Schmitt, M. Loskyll, and D. Zuhlke, "Human-machine-interaction in the industry 4.0 era," in *Proc. 12th IEEE Int. Conf. Ind. Informat. (INDIN)*, Jul. 2014, pp. 289–294.

[52] K. P. Schneider, *Small Talk: Analyzing Phatic Discourse*, vol. 1. Marburg, Germany: Hitzeroth, 1988.

[53] M. Rheu, J. Y. Shin, W. Peng, and J. Huh-Yoo, "Systematic review: Trust-building factors and implications for conversational agent design," *Int. J. Hum. Comput. Interact.*, vol. 37, no. 1, pp. 81–96, Jan. 2021.

[54] L. Clark, N. Pantidi, O. Cooney, P. Doyle, D. Garaialde, J. Edwards, B. Spillane, E. Gilmartin, C. Murad, C. Munteanu, V. Wade, and B. R. Cowan, "What makes a good conversation?: Challenges in designing truly conversational agents," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–12.

[55] S. Tseng and B. J. Fogg, "Credibility and computing technology," *Commun. ACM*, vol. 42, no. 5, pp. 39–44, May 1999.

[56] F. Nothdurft, G. Bertrand, H. Lang, and W. Minker, "Adaptive explanation architecture for maintaining human-computer trust," in *Proc. IEEE 36th Annu. Comput. Softw. Appl. Conf.*, Jul. 2012, pp. 176–184.

[57] A. Glass, D. L. McGuinness, and M. Wolverton, "Toward establishing trust in adaptive agents," in *Proc. 13th Int. Conf. Intell. User Interface (IUI)*, 2008, pp. 227–236.

[58] S. Gulati, S. Sousa, and D. Lamas, "Design, development and evaluation of a human-computer trust scale," *Behav. Inf. Technol.*, vol. 38, no. 10, pp. 1004–1015, Oct. 2019.

[59] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cognit. Ergonom.*, vol. 4, no. 1, pp. 53–71, Mar. 2000.

[60] J. M. Kraus, "Psychological processes in the formation and calibration of trust in automation," Ph.D. dissertation, Dept. Hum. Factors, Universität Ulm, Ulm, Germany, 2020.

[61] M. McTear, "Conversational AI: Dialogue systems, conversational agents, and chatbots," *Synth. Lectures Human Lang. Technol.*, vol. 13, no. 3, pp. 1–251, Oct. 2020.

[62] D. J. Litman and S. Pan, "Designing and evaluating an adaptive spoken dialogue system," *User Model. User-Adapted Interact.*, vol. 12, nos. 2–3, pp. 111–137, 2002.

[63] F. Nothdurft, F. Richter, and W. Minker, "Probabilistic human-computer trust handling," in *Proc. 15th Annu. Meeting Special Interest Group Discourse Dialogue (SIGDIAL)*, 2014, pp. 51–59.

[64] G. Behnke, M. Schiller, M. Kraus, P. Bercher, M. Schmautz, M. Dorna, W. Minker, B. Glimm, and S. Biundo, "Instructing novice users on how to use tools in DIY projects," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 5805–5807.

[65] M. Kraus, M. Schiller, G. Behnke, P. Bercher, S. Biundo, B. Glimm, and W. Minker, "A multimodal dialogue framework for cloud-based companion systems," in *Proc. 9th Int. Workshop Spoken Dialogue Syst. Technol.* Singapore: Springer, 2019, pp. 405–410.

[66] G. Behnke, M. Schiller, M. Kraus, P. Bercher, M. Schmautz, M. Dorna, M. Dambier, W. Minker, B. Glimm, and S. Biundo, "Alice in DIY wonderland or: Instructing novice users on how to use tools in DIY projects," *AI Commun.*, vol. 32, no. 1, pp. 31–57, 2019.

[67] M. Kraus, M. Schiller, G. Behnke, P. Bercher, M. Dorna, M. Dambier, B. Glimm, S. Biundo, and W. Minker, "'Was that successful?' on integrating proactive meta-dialogue in a DIY-assistant using multi-modal cues," in *Proc. Int. Conf. Multimodal Interact.* New York, NY, USA: Association for Computing Machinery, 2020, pp. 585–594, doi: 10.1145/3382507.3418818.

[68] M. Klepsch, F. Schmitz, and T. Seufert, "Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load," *Frontiers Psychol.*, vol. 8, p. 1997, Nov. 2017.

[69] J. Sweller, J. J. G. van Merrienboer, and F. G. W. C. Pass, "Cognitive architecture and instructional design," *Educ. Psychol. Rev.*, vol. 10, no. 3, pp. 251–296, 1998.

[70] B. Laugwitz, M. Schrepp, and T. Held, "Konstruktion eines fragebogens zur messung der user experience von softwareprodukten," in *Mensch und Computer 2006: Mensch und Computer im Strukturwandel*. München, Germany: Oldenbourg Verlag, 2006, pp. 125–134.

[71] B. Rammstedt, C. Kemper, M. C. Klein, C. Beierlein, and A. Kovaleva, "Eine kurze skala zur messung der fünf dimensionen der persönlichkeit: Big-five-inventory-10 (BFI-10)," *Methoden, Daten, Analysen*, vol. 7, no. 2, pp. 233–249, 2013.

[72] S. M. Merritt, H. Heimbaugh, J. Lachapell, and D. Lee, "I trust it, but i don't know why: Effects of implicit attitudes toward automation on trust in an automated system," *Hum. Factors*, vol. 55, no. 3, pp. 520–534, 2013.

[73] T. Nomura, T. Kanda, and T. Suzuki, "Experimental investigation into influence of negative attitudes toward robots on human–robot interaction," *Ai Soc.*, vol. 20, no. 2, pp. 138–150, 2006.

[74] G. V. Glass, P. D. Peckham, and J. R. Sanders, "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance," *Rev. Educ. Res.*, vol. 42, no. 3, pp. 237–288, 1972.

[75] M. Kraus, J. Kraus, M. Baumann, and W. Minker, "Effects of gender stereotypes on trust and likability in spoken human-robot interaction," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, 2018, pp. 1–7.

[76] C. Tannenbaum, R. P. Ellis, F. Eyssel, J. Zou, and L. Schiebinger, "Sex and gender analysis improves science and engineering," *Nature*, vol. 575, no. 7781, pp. 137–146, Nov. 2019.

[77] T. Law, M. Chita-Tegmark, and M. Scheutz, "The interplay between emotional intelligence, trust, and gender in human–robot interaction," *Int. J. Social Robot.*, vol. 13, no. 2, pp. 297–309, 2020.

[78] S. E. Seibert, J. M. Crant, and M. L. Kraimer, "Proactive personality and career success," *J. Appl. Psychol.*, vol. 84, no. 3, p. 416, 1999.

MATTHIAS KRAUS was born in Lauingen, Bavaria, Germany, in 1990. He received the B.Sc. degree in electrical engineering from Ulm University, Germany, in 2014, and the M.Sc. degree in communications and computer engineering with a focus on human–machine interaction, in 2016. He is currently pursuing the Ph.D. degree in communications and computer engineering with Ulm University. Since 2017, he has been a Research Assistant with the Dialogue Systems Group, Ulm University. His research interest includes the development of situation- and user-based proactive interaction strategies in multi-modal dialogue systems.

NICOLAS WAGNER was born in Marktoberdorf, Germany, in 1992. He received the B.Sc. and M.Sc. degrees in communications and computer engineering with a focus on human–machine interaction, in 2016 and 2018, respectively. He is currently pursuing the joint Ph.D. degree with the University of Granada, Spain. Subsequently, he joined the Dialogue Systems Group, Ulm University, Germany, as a Research Assistant. His research interests include multi-user dialogue systems and human–robot interaction.

ZORAIDA CALLEJAS (Member, IEEE) received the Ph.D. degree in computer science from the University of Granada, Spain, in 2008. She has been a Visiting Researcher at the Technical University of Liberec, Czech Republic; the University of Trento, Italy; Ulster University, U.K.; the Technical University of Berlin, Germany; Ulm University, Germany; and Télécom ParisTech, France. She is currently a Tenured Associate Professor with the Department of Software Engineering, University of Granada. She currently coordinates the EU H2020 MSCA-RISE Project MENHIR about the use of conversational technologies to promote mental health. She is also coordinating the Andalusian I + d + i Project BonAppPetit related to the use of conversational systems to foster healthy eating in children. Her research interests include dialogue systems, conversational systems, emotion processing, affective computing, and user modeling.

WOLFGANG MINKER received the Diploma (M.Sc.) degree in engineering science from the University of Karlsruhe, Germany, the Ph.D. degree in engineering science from the University of Karlsruhe, in 1997, and the Ph.D. degree in computer science from Université Paris-Sud, France, in 1998. From 1993 to 2000, he was a Teaching Assistant at LIMSI-CNRS, Université Paris-Sud, and subsequently worked as a Senior Researcher at the Dialogue Systems Group, DaimlerChrysler Research and Technology, Ulm, Germany, from 2000 to 2003. Since 2003, he has been a Full Professor with Ulm University, Ulm. He became the Co-Director of the International Research Laboratory Multimodal Biometric and Speech Systems, ITMO University, St. Petersburg, Russia, in 2017.

• • •