User Involvement in Collaborative Decision-Making Dialog Systems

Florian Nothdurft, Pascal Bercher, Gregor Behnke and Wolfgang Minker

Abstract Mixed-initiative assistants are systems that support humans in their decision-making and problem-solving capabilities in a collaborative manner. Such systems have to integrate various artificial intelligence capabilities, such as knowledge representation, problem solving and planning, learning, discourse and dialog, and human-computer interaction. These systems aim at solving a given problem autonomously for the user, yet involve the user into the planning process for a collaborative decision-making, to respect e.g. user preferences. However, how the user is involved into the planning can be framed in various ways, using different involvement strategies, varying e.g. in their degree of user freedom. Hence, here we present results of a study examining the effects of different user involvement strategies on the user experience in a mixed-initiative system.

Keywords Human-computer interaction · Cooperative decision-making · User experience · Dialogue systems

F. Nothdurft (✉) · W. Minker
Institute of Communications Engineering, Ulm University, Ulm, Germany
e-mail: florian.nothdurft@uni-ulm.de

W. Minker
e-mail: wolfgang.minker@uni-ulm.de

P. Bercher · G. Behnke
Institute of Artificial Intelligence, Ulm University, Ulm, Germany
e-mail: pascal.bercher@uni-ulm.de

G. Behnke
e-mail: gregor.behnke@uni-ulm.de

1 Introduction

Most contemporary dialog systems (DS) manage the interaction between human and machine in a uni-directional dependency. Most common, users interact with a DS to solve domain-dependent tasks. However, this is usually limited to information retrieval or exchange tasks, such as searching for bus connections or restaurants, where usually the user is exclusively in charge. Contrary to that, future DS will evolve towards being *companions* [1] for the user (i.e. intelligent personal assistants) that help the user not only in simple but also in complex tasks. These companions solve complex problems collaboratively with the user, if either the user or the DS is not able to solve the problem on its own, to the liking of the other, or simply if the user's load should be reduced. Companions may, for example, provide assistance in form of artificial intelligence problem solving skills (i.e. planning), that are intrinsically different from human problem solving skills. AI planning may solve combinatory problems that are too complex for humans, due to the need of mathematical computations or fast information storage and retrieval of large data. In addition, such systems may have knowledge and problem solving skills for a domain, the user is not an expert in.

However, the automatic and autonomous generation of a solution (i.e., a plan) to a user-provided problem by such artificial skills potentially does not respect the user's individual needs and preferences, and is per se not always the best solution to the problem. The generated plan is usually only *a* solution, and not one which is best-suited for the user. Integrating preferences into planning [2] is a solution, yet requires the user to specify his preferences a priori in an expressible (e.g. action or method costs) way, which is likely to result in user frustration or even interaction stop. In addition, the planning process is done autonomously and exclusively by the planning system. However, this exclusion of the user from the decision-making process will lead to a couple of problems, we described in previous work [3]. For example, if humans are not involved into a decision-making process, they are less likely to follow or execute a proposed plan or solution. In addition, in decisions that involve grave risks, e.g. in military settings [4] or spaceflight [5], humans must have the final decision on which actions are to be contained in the plan.

Therefore, we proposed a collaborative decision-making assistant, in Nothdurft et al. [3], that combines AI planning and human problem-solving skills into a collaborative decision-making process. This results in a mixed-initiative planning system (MIP) [4–6], or more general a mixed-initiative assistant (MIA) [7] that supports users in problem solving and finding appropriate solutions. A collaborative decision-making process has the intent of solving a problem the user is not able to solve at all or only with great effort. It aims at relieving the user's cognitive load and simplifying the problem at hand. In general, the intertwining of human and AI decision-making skills should lead to an increased user-experience by more preferable and individual solutions for the user. In addition, MIP also facilitates the adaptation of a companion to its owner. The companion may learn from past interaction episodes and direct the future decision-making processes to the user's liking. Since not only the companion

may adapt to the user over time, but the user to the decision-making capabilities of the system as well, this process may be described as a *co-adaptation* of two parties.

However, intertwining user and AI planning systems into a MIP system, does not only facilitate more intelligent and competent systems, but does also raise new questions. In previous work we described the potentials, challenges, and risks involved in such MIP systems, along a prototypical MIP system architecture. Some of these challenges, for example, how to maintain coherent models for the participating components [8], or how to deal with occuring phenomena, such as *backtracking* in a collaborative MIP process [3], were already tackled.

2 Mixed-Initiative Planning

In general, the interaction between AI planning and user has to begin with a dialog to define the statement of an objective. This first dialog has the goal of defining the task in a way understandable for the planner. Once the problem is passed to the planner the interactive planning itself may start. Using a selected search strategy (here: *depth-first search*) the plan is refined by selecting appropriate modifications for open decisions. In order to decide whether to involve the user or not during this process, an elaborate decision model, integrating various information sources, is used. Relevant information sources are, e.g. the *dialog history* (e.g. was the user's decision the same for all past similar episodes?), the kind of *open plan decision* (e.g. is this decision relevant for the user?), the *user profile* (e.g. does the user have the competencies for this decision?), or the current *situation* (e.g. is the current cognitive load of the user low enough for interaction?). These information sources are used in a superordinate component, the *decision model*, to decide whether to involve the user. The decision model can either initiate a user interaction or determine by itself that the planner should make the decision. This is equivalent with the user signaling no explicit preference in the decision-making. Furthermore, it is important whether the additional interaction is critical and required, to successfully continue the dialog. Additional dialogs may contribute to achieving short-term goals, but risk the user's cooperativeness, in the long run, e. g. by overstraining his cognitive capabilities or boring him.

In case of user involvement the information on the current *plan decision* has to be communicated to the user. This means that the open decision and the corresponding choice between available *modifications* have to be represented in a dialog suitable for the user. Hence, the corresponding plan actions or methods (i.e. the set of possible actions for a upcoming decision) needs to be mapped to human-understandable dialog information. As this mapping is potentially required for every plan action or method, and vice versa for every dialog information, coherent models between planner and DS become crucial for MIP systems. The thorough matching of both models would be an intricate and strenuous process, requiring constant maintenance, especially when models need to be updated. Thus, a more appropriate approach is the automatic generation or respectively extension of the respective models using one
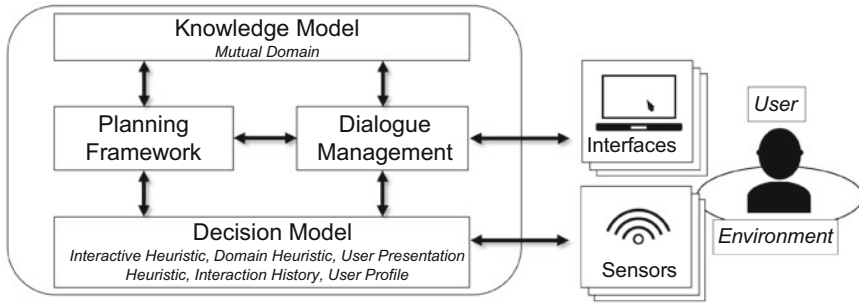
Fig. 1 Essential components of a MIP system (cf. [3])

mutual model as source, the *mutual knowledge model* (cf. [8]). From this model—in this case an OWL ontology [9]—the dialogs and their hierarchy can be derived, using the topmost elements as entry points for the dialog between user and machine. This is, for example, needed for the user to specify the objective for the planner, or to present the available *plan modifications* (i.e. the options in the decision-making), that have to be translated to a format understandable to the user (cf. [3]). The model is also used to extend the planning domain: hierarchical structures (i.e. decomposition methods) are derived using declarative background knowledge modeled in the ontology (cf. [8]). Using a mutual model addresses one of the challenges of MIP (cf. [3]), since translation problems between dialog and planner semantics can be prevented, even when updating the domain (e.g. by acquiring new knowledge, such as new workouts or rehabilitation methods). Another challenge related to the specific interaction between man and machine is *if*, *how*, and at *what* specific point in the dialog user-involvement is necessary or useful. This is one of the most essential challenges, as the integration process, and how the shift of initiative towards one of the parties is framed, affects how effective and user-friendly the MIP will be (Fig. 1).

Our multimodal MIP system was implemented using a knowledge-based cognitive architecture. The multimodal interface uses mainly speech, and graphical user interfaces as input and output. In addition, gestures, and sensory information such as the user location can be used as input. The use of multiple modalities enables us to vary the means of collaboration from uni- to multimodal interaction. The *Dialog Management* uses a modality-independent representation, communicating with the user via the *Fission* [10], *User Interface* [11], *Sensors* [12], and *Fusion* [13] modules.

3   Related Work

Initial work on combing dialog and planning in a mixed-initiative fashion has been done by George Ferguson and James Allen in their TRAINS [14] and TRIPS [15] systems. Their systems include the collaborative capabilities of reasoning, planning,

execution, and communication and are based on the belief-desire-intention (BDI) model of agency [16]. Important work, approaching the problem from a different perspective, has been done by Rich et al. in COLLAGEN [17], aiming at applying collaborative discourse theory to human-computer interaction. Their work is based on the SharedPlans theory [18], and models the dialog state of the agents (i.e. user and system) as they interact and perform activities. More recent work involving mixed-initiative interaction has been done in various application domains (e.g. [4, 5, 19, 20]). One of the most well known is MAPGEN [5], applying a mixed-initiative planning and scheduling approach for the ground operations system for the Mars Exploration Rover of NASA. Abstract goals were planned by the user, yet the planner assured that all constraints, which is very complex in such a setting, are satisfied. Another example is DiamondHelp [21], a generic collaborative task-guidance system, which may also integrate the COLLAGEN system. DiamondHelp can be used for a multitude of tasks (e.g. help the user in programming a washing machine or thermostat).

What these work is missing is to investigate *how* the user's involvement should be framed. If the user is to be involved, the question arises how this should be rendered, i.e. what kind of integration is the most beneficial. In addition, if the user is not involved in the decision-making, it has to be decided if and how the user may be informed about the decisions the planner has made. The decision whether and how to involve the user into the planning process is not only controlled by a degree of necessity dependent on the current task and situation, but should also take into consideration the effects on the user's system experience. Usually, the user involvement is done by presenting a list of possible options for upcoming decisions to the user. If this form of user involvement is always necessary or simply best for the user experience is rather questionable. User involvement strategies may actually range from almost unrestricted decision-making (i.e. set of options), limited only by valid solution constraints, over explicit confirmations of system-preselected decisions, to only informing the user of made decisions. Hence, we designed the study examining the effects of different strategies of user involvement on the user experience.

4 User Study About User Involvement Strategies

For this study, we used our prototypical MIP system [3] and implemented several strategies to involve the user into the decision-making. This means that we evaluated different degrees of user involvement into a planning process, ranging from only informing the user of system-made decisions to explicitly requesting a user confirmation for the proposed system decision. In this scenario the user's task was to create individual strength training workouts. In each strength training workout at least three different muscle groups had to be trained and exercises chosen accordingly. The user was guided through the process by the system, which provided a selection of exercises for training each specific muscle group necessary for the workout. For example, when planning a strength training for the upper body, the user had to select exercises

to train the chest. This selection is an involvement of the user into the MIP process. The decision how to refine the task of training the chest is not made by the system, but left to the user. The system decision was based on previously made selections by the user. This means that when in a previous interaction the same decision (i.e. the same situation with the exact same options) had do be done, this user-selected option was remembered for future interactions, and selected accordingly by the system. Of course, in a more complex scenario this decision would depend not only on the interaction history, but also on additional information (e.g. affective user states like overextension, interest, or engagement) stored in the user state. The system-made selection was presented in various ways, which were the following:

Explicit confirmation (EC)   based on previous selections the choice was already made by the system and presented to the user, who had to explicitly confirm the choice by clicking "okay".

Implicit confirmation (IC)   the system-made decision (i.e. the selection) was presented to the user, but the user could intervene, in a certain time frame, by clicking "Let me decide". Therefore, this is a form of implicit confirmation.

Information (INF)   the system-made decision was presented to the user without the need of confirmation. Hence, the user was only informed of the system's decision-making, without the option to intervene.

Unsorted (US)   the baseline was the usual unsorted selection task. No proactive behavior by the system was present, meaning that users had to select from a list.

In all conditions the system-made selection was explained by the system using a phrase similar to "For training this muscle group, you previously selected this exercise. Therefore, it was already selected for you." The participants were distributed by a random-function to the variants, resulting in 23 participants receiving the known unsorted selection, 25 asked for explicit confirmation, 30 with implicit confirmations, and 26 receiving only an information by the system.

4.1   Used Questionnaires

For the assessment of the study we chose two questionnaires. The AttrakDiff 2 questionnaire [22], which extends the assessment of technical systems or software in general from the limited view of usability, which represents mostly pragmatic qualities, to the integration of scales measuring hedonic qualities. It consists of four basic scales: *perceived pragmatic quality*, which measures the product's ability to achieve the user's goals efficiently and effectively without inducing a high mental load; *hedonic quality-stimulation*, which measures whether novel, interesting and inspiring qualities are present to increase the user's attention and foster the user's abilities and skills; *hedonic quality-identity*, which assesses the user's perceived identity of the subject at evaluation; and *perceived attractiveness*, which is a global rating based on the perceived qualities (Fig. 2).
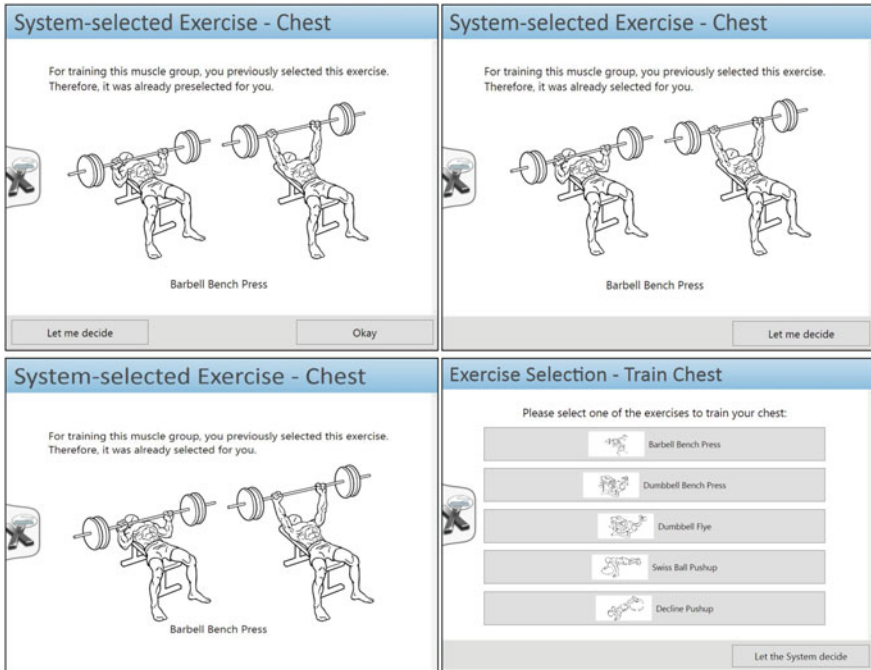
Fig. 2 The different degrees of user involvement: on the *top left* the proactive decision by the system has to be confirmed explicitly by the user (EC). On the *top right*, the decision is presented and the user may intervene in a given time frame, else it is confirmed (IC). On the *bottom left* the user is only informed of the system's decision-making (INF), and on the *bottom right* the usual selection is presented as baseline condition (US)

The other used questionnaire measured the cognitive load. Cognitive load, which consists of the three different types, should not exceed the working memory capacity [23]. One of the basic ideas of cognitive load theory is that a low extraneous load, resulting from a good instructional design, enhances the potential that users engage in cognitive processes (i.e. germane load) related to learning [24]. Hence, the better the instructional design, the greater potential for germane cognitive load and learning. We used an experimental questionnaire developed by [25] which measures all three types of cognitive load separately. The questionnaire consisted of 12 items, with four items each for every type of cognitive load: *intrinsic cognitive load*, which can be described as the inherent load induced by the content itself. This type of load can not be changed by the design of the learning material and is caused mainly by the difficulty of the task. In other words it results e.g. from the number of elements that must be simultaneously processed in the working memory; *extraneous cognitive load*, which is caused by the presentation form of the learning material and is considered to be manipulable by the design of the learning material; *germane cognitive load*, which is considered the load inflicted by the learning process. Germane cognitive load is "good" cognitive load, which helps in fostering the processes inherent in the construction and automation of schemas.

4.2 Hypotheses

Our hypotheses were that in general the various conditions will perform differently, especially regarding perceived cognitive load, pragmatic qualities and attractiveness of the system. We expected no significantly different influences on the human-computer trust relationship between human and machine for the conditions. The exclusion of the user from the decision making (i.e. only informing the user) was expected to reduce the hedonic quality compared to the use of explicit and implicit confirmations. The baseline was expected to perform worst for the perceived pragmatic system quality, and the explicit confirmation best. In terms of cognitive load we expected that when the system takes over the decision making (i.e. implicit confirmation or user information), the cognitive load for the user is reduced compared to the other conditions.

4.3 Results

The results were collected using the AttrakDiff, cognitive load, and human-computer trust questionnaire. In addition, we used an open questions form for user feedback. As the conditions would not affect objective measures like task completion or efficiency rate, they were neglected in this paper.

AttrakDiff Assessing the results of the AttrakDiff questionnaire, using a one-way ANOVA, we found marginal differences between the conditions for the dimensions (see Table 1) of perceived *hedonic quality-identity* ($F(3, 96) = 2.172$, $p = 0.096$) and the perceived overall *attractiveness* ($F(3, 96) = 2.420$, $p = 0.071$) of the system. Post-hoc comparisons using the Fisher Least significant difference (LSD) test indicated that the mean score of *hedonic quality-identity* for the US condition ($M = 3.71$, $SD = 0.705$) was significantly different, at the $p = 0.015$ level, than the INF condition ($M = 4.37$, $SD = 0.86$). For *attractiveness* the US condition ($M = 3.88$, $SD = 0.77$) was also significantly different ($p = 0.009$) than the INF condition ($M = 4.62$, $SD = 0.93$).

Table 1 This table shows the mean values of the AttrakDiff questionnaire dimensions

|		PQ	HQ-I	HQ-S	ATT
US	M	4.27	3.71	3.77	3.88
	SD	0.91	0.71	1.02	0.77
EC	M	4.47	4.13	3.95	4.17
	SD	1.12	1.29	1.17	1.24
IC	M	4.31	3.97	3.97	4.19
	SD	1.12	0.68	0.77	0.84
INF	M	4.81	4.37	4.10	4.62
	SD	0.99	0.87	0.96	0.93
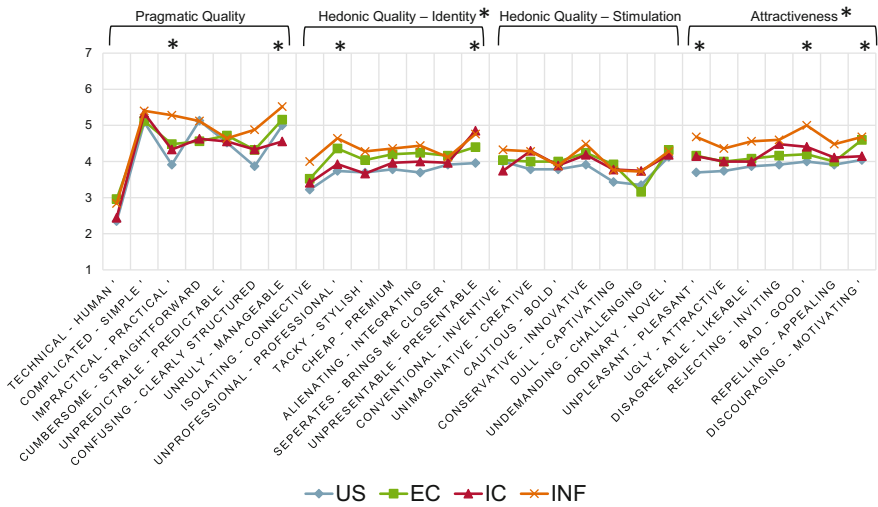
Fig. 3 This shows the average means of the AttrakDiff comparing the confirmation conditions on a 7-point Likert scale. US is a unsorted list of options, EC and IC explicit and implicit confirmations, and INF only informs the user of the system decision. The ' indicates inverted, for the sake of readability, scales. The * indicates significance

Looking further into the data and analyzing the single word pairs of the AttrakDiff questionnaire (see Fig. 3), to find the origin of the differences, we could find more detailed results. Using one-way ANOVA tests we found significant differences in the word pair *impractical-practical* and marginal significance for *unruly-manageable*, which both belong to the dimension of *pragmatic quality*. Post hoc comparisons using Fishers LSD showed that for *impractical-practical* the INF condition performed significantly better ($p = 0.002$) than the US condition and also significantly better ($p = 0.024$) than the IC condition. For *unruly-manageable* the INF condition performed significantly better ($p = 0.14$) than the IC condition.

For the dimension of *hedonic quality-identity* we found a marginal significant differences for the word pairs *unprofessional-professional* and *unpresentable-presentable*. Fishers LSD post hoc test revealed that INF performed significantly better ($p = 0.018$) than US for *unprofessional-professional*. For *unpresentable-presentable* the IC condition was significantly better ($p = 0.020$) than US, which was also significantly worse ($p = 0.040$) than the INF condition.

For *attractiveness* a marginal significant difference, using also a one-way ANOVA, was found in *unpleasant-pleasant* ($F(3, 96) = 2.211, p = 0.092$), *bad-good* ($F(3, 96) = 2.397, p = 0.073$), and *discouraging-motivating* ($F(3, 96) = 2.314, P = 0.081$). Post hoc tests revealed that the significant differences for *unpleasant-pleasant* were between INF and US at ($p = 0.012$). For *bad-good* the average mean of the INF condition was significantly better at ($p = 0.014$) than US, and also better at ($p = 0.043$) than the EC condition The third word pair showing significant
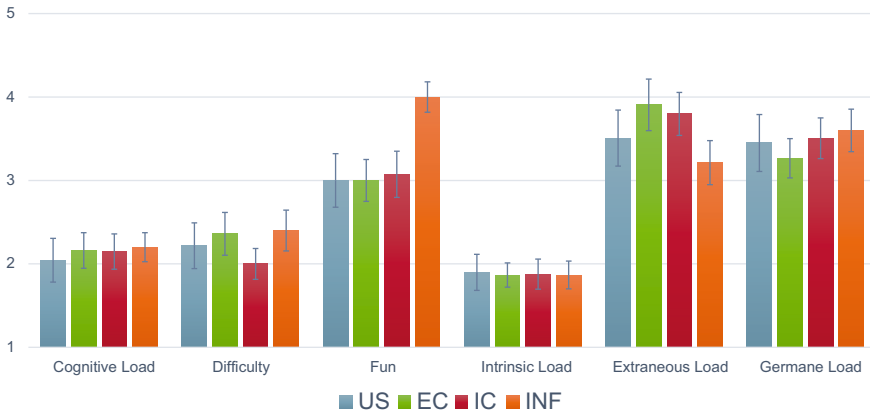
Fig. 4 This shows the average means of the cognitive load comparing the confirmation conditions on a 7-point Likert scale. Intrinsic, extraneous and germane load originate from the experimental questionnaire

results ($p = 0.037$) was *discouraging-motivating* with INF performing better than US. In addition, we found also that the user information condition was significantly performing better for *ugly-attractive* with US and INF at ($p = 0.037$), for *rejecting-inviting* with US and INF at ($p = 0.039$), and also for *repelling-appealing* with US ($M = 3.91, SD = 1.08$) and INF at ($p = 0.046$).

Cognitive Load Analyzing the cognitive load items (see Fig. 4) we found significant differences, using a one-way ANOVA, for *fun* with ($F(3, 96) = 3.488, p = 0.019$). Fishers LSD showed that the user information condition ($M = 4.00, SD = 0.91$) was significantly better than the rest. Compared to US ($M = 3.00, SD = 1.53$) at ($p = 0.009$), to EC ($M = 3.00, SD = 1.25$) at ($p = 0.008$), and to IC ($M = 3.07, SD = 1.43$) it was significant at ($p = 0.012$).

Open Questions The following comments were entered by the participants: 'carrying over previous made decisions should be confirmed by the user', and 'If the system selects an exercise, the system should notify why this exercise was thought to be the most fitting one'.

4.4 Discussion

Surprisingly it showed that only informing the user of the system-made selection, without any possibility to intervene, was performing best in almost any category. The pragmatic quality, the identification with the system (hedonic quality) and the overall attractiveness were best for the INF condition. The automatic selection of the system was perceived as very practical and increased the perception that the system is predictable and manageable. This goes along with the fact that the system

behavior was explained to the user, considering earlier results on explanations and system acceptance. Additionally, the INF condition was experienced as the most enjoyable of all, along with reducing the extraneous load (cf. Fig. 4) of the task at hand. Even though the baseline condition of selecting from an unsorted list as before, was experienced before, and thus would require no additional cognitive load, the automatic selection and informing the user of this decision tends to be less demanding on the extraneous load. Also the technical competence of the system was perceived better than for the baseline condition.

The integration of the user into the decision making using explicit confirmations seem to perform second best for most dimensions and items. Though it seems to increase the extraneous load of the task, by requiring additional user input, the identification of the user with the system, measured by the *hedonic quality-identity* seems to be greater. The fact that the *implicit confirmation* condition performed that much worse than the *user information* actually seems odd to us. It appears to us that the combination of informing the user and presenting, for a defined time frame, the explicit interaction possibility to deny the automated selection, was confusing for the user. Maybe the button labeled 'Let me decide', or the definition of a restricted time frame was not clearly understandable, thus leading to a worse user-experience. These results show that the decision on how to frame the interaction dialog between user and system will affect the user experience and potentially the cognitive load of the user. As future DS will become more complex and evolve to collaborative intelligent assistants rather than simple problem solvers, the way of interaction between those two parties will be crucial.

5 Conclusion

Overall, it seems, that for decisions which are understandable and reasonable, informing the user of system-made decisions may contribute to a more practical, attractive, fun and less demanding dialog system. However, one must be careful to transfer these findings to other domains or more complex tasks. The positive experience of the user information condition might be due to the task at hand. Usually, workouts are planned, at least for inexperienced users, by experts (e.g. coaches). Addressing competences to a workout planner system, and therefore trusting its decisions, seems like a logical conclusion. For future evaluations it will be interesting to compare these results to automated system behavior for tasks, where usually the user is in charge and dictates the decision-making process. This might lead to a decrease of acceptance for user information conditions and to an increase for conditions, where the user has more control. Nevertheless, this work shows that it is important to investigate in the collaboration dialog (e.g. user involvement strategies) between user and system, which will be important for future more intelligent and assistive dialog systems.

References

1. Wendemuth, A., Biundo, S.: A companion technology for cognitive technical systems. In: Anna Esposito, Alessandro Vinciarelli, R.H.V.C.M. (ed.) Proceedings of the EUCogII-SSPNET-COST2102 International Conference, pp. 89-103 (2011). Lecture Notes in Computer Science, Springer Berlin Heidelberg (2012)

2. Sohrabi, S., Baier, J., McIlraith, S.A.: HTN planning with preferences. In: Proceedings of the 21st Int. Joint Conference on Artificial Intelligence (IJCAI 2009). pp. 1790-1797. AAAI Press (2009)

3. Nothdurft, F., Behnke, G., Bercher, P., Biundo, S., Minker, W.: The interplay of user-centered dialog systems and ai planning. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL). pp. 344-353. Association for Computational Linguistics, Prague, Czech Republic (September 2015)

4. Myers, K.L., Tyson, W.M., Wolverton, M.J., Jarvis, P.A., Lee, T.J., desJardins, M.: PASSAT: a user-centric planning framework. In: Proceedings of the 3rd International NASA Workshop on Planning and Scheduling for Space, pp. 1-10 (2002)

5. Ai-Chang, M., Bresina, J., Charest, L., Chase, A., Hsu, J.J., Jonsson, A., Kanefsky, B., Morris, P., Rajan, K., Yglesias, J., et al.: Mapgen: mixed-initiative planning and scheduling for the mars exploration rover mission. IEEE Intell. Syst. **19**(1), 8–12 (2004)

6. Fernández-Olivares, J., Castillo, L.A., García-Pérez, Ó., Palao, F.: Bringing users and planning technology together. Experiences in SIADEX. In: Proceedings of the 16th International Conference on Automated Planning and Scheduling (ICAPS 2006). pp. 11-20. AAAI Press (2006)

7. Tecuci, G., Boicu, M., Cox, M.T.: Seven aspects of mixed-initiative reasoning: an introduction to this special issue on mixed-initiative assistants. AI Mag. **28**(2), 11 (2007)

8. Behnke, G., Ponomaryov, D., Schiller, M., Bercher, P., Nothdurft, F., Glimm, B., Biundo, S.: Coherence across components in cognitive systems–one ontology to rule them all. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2015). AAAI Press (2015)

9. W3C OWL Working Group: OWL 2 Web Ontology Language: Document Overview (2009). http://www.w3.org/TR/owl2-overview/

10. Honold, F., Schüssel, F., Weber, M.: Adaptive probabilistic fission for multimodal systems. In: Proceedings of the 24th Australian Computer-Human Interaction Conference, pp. 222-231. OzCHI'12, ACM, New York, NY, USA (November, 26-30 2012)

11. Honold, F., Schussel, F., Weber, M., Nothdurft, F., Bertrand, G., Minker, W.: Context models for adaptive dialogs and multimodal interaction. In: Proceedings of the 9th International Conference on Intelligent Environments (IE 2013), pp. 57-64. IEEE (2013)

12. Glodek, M., Honold, F., Geier, T., Krell, G., Nothdurft, F., Reuter, S., Schüssel, F., Hoernle, T., Dietmayer, K., Minker, W., Biundo, S., Weber, M., Palm, G., Schwenker, F.: Fusion paradigms in cognitive technical systems for Human-Computer interaction. Neurocomputing **161**, 17–37 (2015)

13. Schüssel, F., Honold, F., Weber, M.: Using the transferable belief model for multimodal input fusion in companion systems. In: Multimodal Pattern Recognition of Social Signals in HCI, LNCS, vol. 7742, pp. 100-115. Springer (2013)

14. Allen, J.F., Schubert, L.K., Ferguson, G., Heeman, P., Hwang, C.H., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M., et al.: The trains project: a case study in building a conversational planning agent. J. Exp. Theor. Artif. Intell. **7**(1), 7–48 (1995)

15. Ferguson, G., Allen, J.F., et al.: Trips: An integrated intelligent problem-solving assistant. In: Proceedings of the AAAI/IAAI, pp. 567–572 (1998)

16. Rao, A.S., Georgeff, M.P.: Modeling rational agents within a bdi-architecture. KR **91**, 473–484 (1991)

17. Rich, C., Sidner, C.L.: Collagen: a collaboration manager for software interface agents. User Model. User-Adapt. Interact. **8**(3–4), 315–350 (1998)

18. Grosz, B.J., Kraus, S.: Collaborative plans for complex group action. Artif. Intell. **86**(2), 269–357 (1996)
19. Myers, K.L., Jarvis, P.A., Tyson, W.M., Wolverton, M.J.: A mixed-initiative framework for robust plan sketching. In: Proceedings of the 13th International Conference on Automated Planning and Scheduling (ICAPS 2003). pp. 256-266. AAAI Press (2003)
20. de la Asunción, M., Castillo, L., Fdez-Olivares, J., García-Pérez, Ó., González, A., Palao, F.: Siadex: an interactive knowledge-based planner for decision support in forest fire fighting. AI Commun. **18**(4), 257 (2005)
21. Rich, C., Sidner, C.L.: Diamondhelp: a generic collaborative task guidance system. AI Mag. **28**(2), 33 (2007)
22. Hassenzahl, M., Burmester, M., Koller, F.: Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualitt. In: Szwillus, G., Ziegler, J. (eds.) Mensch & Computer 2003: Interaktion in Bewegung, pp. 187–196. B. G. Teubner, Stuttgart (2003)
23. Kirschner, P.A.: Cognitive load theory: implications of cognitive load theory on the design of learning. Learn. Instr. **12**(1), 1–10 (2002)
24. Paas, F., Van Merriënboer, J.: Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. J. Educ. Psychol. **86**(1), 122 (1994)
25. M. Klepsch, F.W., Seufert, T.: Differentiated Measurement of Cognitive Load: Possible or Not? (2015), in preparation