How Developers Anticipate User Behavior in the Design of Assistance Systems

Cordula Krinner

Technische Universität Berlin
Prospective Design of Human-Technology Interaction Group
Franklinstr. 28–29, 10587 Berlin, Germany
Cordula.Krinner@zmms.tu-berlin.de

Abstract. This paper proposes a new perspective on the old problem of function allocation. Instead of analyzing a synchronous interaction between human and computer, we suggest it could be more helpful to reconceptualize function allocation as an asynchronous division of labor between developers/designers and operators/users of human-computer systems. A study with 31 teams of developers was conducted in order to determine the effect of various forms of contact with a system and user participation on the results of a development process. The implications of lacking expected effects for future research are discussed.

Keywords: Assistance systems, allocation of functions, developers, participative design.

1 Introduction

The problem of allocating tasks in human-technology systems seems to be a never-ending story. Fitts provides one of the oldest propositions to tackle that problem, dating back to 1951 [1]. Although following this, many and apparently novel ways of dealing with this problem have been proposed by various authors, Dekker and Woods [2] claim in an article that even some of today's research and literature on function allocation is still pursuing an "Abracadabra dream of MABA-MABA methods: put your allocation problem into our method, and the solution will emerge from the other end" [2]. Instead of pursuing the silver bullet of MABA-MABA, this paper tries to shed a new light on the problem of functional allocation in human-technology systems. We propose to reconceptualize human-technology interaction as a division of labor between developers and operators of systems.

2 Theoretical Background

2.1 Function Allocation

The term function allocation is used to describe the question which functions in a human-computer or human-technology system in general should be assigned to

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 98–107, 2007.

humans, e. g. operators or users and which to machines or computer programs. Fitts wrote a report with one of the first approaches to that question [1]. It included a listing of respective abilities of humans and machines, which is referred to as "Fitts' list" or "MABA-MABA"-list, a short form for "Men Are Better At – Machines Are Better At". The question whether humans or technological devices are superior in which area was not only addressed by Fitts and other list accounts. Sheridan and Verplank developed a taxonomy of different levels of automation [3]. An enhanced version of this taxonomy taking into account stages of human action was proposed by Parasuraman, Sheridan, and Wickens [4]. Kaber and Endsley described another way to classify automation [5]. Maybe a more philanthropic approach to classification of human-technology cooperation consists in categorizing technical *assistance* to human actions instead of *automation* as a substitute for human action [6,7]. Technical assistance can be conceptualized as a subdivision of technical automation; yet the classifications for automation mentioned above do not cover all aspects that are relevant for assistance functions or systems.

In the following paragraph, a short description of Wandke's taxonomy will be given [7]. He proposes classifying assistance systems on a total of five dimensions. The central dimension is stage of human action which is supported by an assistance system. This dimension can be further subdivided into six stages that constitute one complete action cycle:

1. motivation, activation, and goal setting
2. perception
3. information integration, generating situation awareness
4. decision making, action selection, and action execution
5. action execution
6. processing feedback of action results.

A second dimension are adjustment possibilities of an assistance system, spanning from fixed systems over adaptable to adaptive systems. The third dimension describes if an assistance system initiates supporting actions, i. e. proactive assistance, or if the user has to ask for support, i. e. reactive assistance. Dimension four describes the possibilities the user has to enter data, e. g. mono-, multimodal or no explicit input. Finally, dimension five comprehends the output of the system, e. g. mono-, multimedia or implicit presentation.

Three other approaches to the problem include designing *for* the user by following principles of user-centered design [8] and designing *with* the user, i. e. applying participative design methods [9,10]. Design might also be carried out *by* the user in that predefined components of a product can be arranged in a way suitable to the individual user. These techniques might at first seem quite distinct from list and various classification approaches, but they try to ensure optimal system performance and optimal function allocation by taking into account characteristics of (future) users.

2.2 Division of Labor Between Developers and Operators

This paper wants to open a new perspective on topics of function allocation and the cooperation between human and technology. Focus of the aforementioned theoretical accounts is a synchronous interaction between human and computer/technology. This view shall be shifted towards an asynchronous division of labor between two groups of people, namely

- those anticipating and planning systems, i. e. developers, programmers or designers, and
- those finally using the implemented systems, e. g. operators or users.

The part of the project described in this paper aims at determining a profile of resources used and contributions made by developers. In collaboration with research conducted concerning operators, it will be possible to compare resources used and special contributions made by each of these two groups.

3 Experimental Paradigm and Research Question

Using the same experimental paradigm throughout all studies, it is possible to compare performances of these two groups. The paradigm consists of a *cooperative* tracking task with decision situations and takes place in a microworld [11,12,13], cf. Fig. 1. An operator's task is to provide on-line assistance and guidance to two individuals fulfilling a simple motor tracking task cooperatively. In contrast, a developer's task is to plan in advance which technical assistance functions might be useful for the cooperatively tracking subjects (to simplify this text, those two subjects will from now on be called "microworld inhabitants" as they are conceptualized as an animate part of the microworld). Implemented assistance systems will later on assist microworld inhabitants performing the

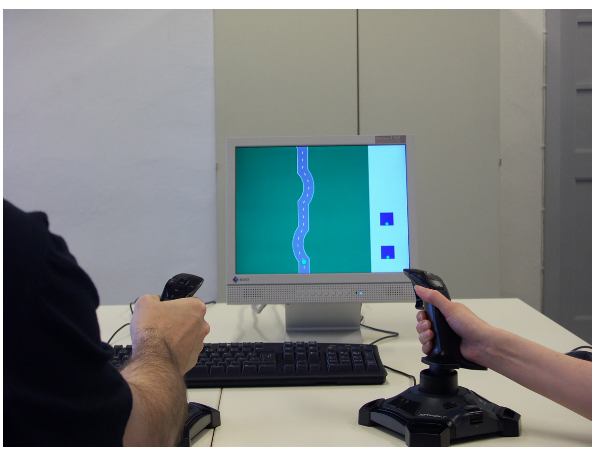

Fig. 1. The cooperative tracking task

cooperative tracking task. Thus, it is possible to compare performance under conditions of real time assistance given by an operator and technical assistance provided by a developer.

The study reported here deals with the research question how one special resource influences process and results of the development of assistance systems. The resource of interest is contact with the system and its users for whom assistance shall be created. It is hypothesized that a greater amount of contact induces more knowledge, which in turn results in better development processes and results. Thus, this study constitutes an experimental check of the tacit agreement that participative design methods [9,10] lead to systems that are more usable. This paper only reports effects of different kinds of exposure on the *results* of developing. Thus, the effects of the independent factor on one general dependent variable—kind of assistance system created—will be reported.

4 Method

4.1 Participants

62 subjects participated; as they developed assistance systems, the terms developers, subjects and participants will be used interchangeably. Subjects were advanced students of technical courses of study, e. g. electrical and mechanical engineering, information science and transportation science. Reflecting the distribution in these courses of study, 49 participants were male, 13 were female. As participants were assigned to teams of two, the relevant sample size is 31 developer teams.

4.2 Design

An experimental design with between-subjects manipulations was chosen. The manipulated factor was varied according to the three following levels:

1. Developers only read a detailed description of characteristics of the system for which they had to develop an assistance system (baseline, condition 1). This description was also given to developers in the other conditions.
2. After reading the description, they performed a cooperative tracking task themselves (condition 2).
3. They watched a screen recording of a cooperative tracking situation. In this video, typical problems of cooperative tracking were exemplified. Cooperative tracking in the screen recording was executed on a predefined route. It was identical to the route on which developers in condition two performed the tracking task on. Furthermore, teams had the opportunity to ask the microworld inhabitants questions, which were answered by the experimenter (condition 3).

Developers' task was to develop one or more concepts to assist cooperative tracking. The dependent variable was kind of assistance system developed.

4.3 Procedure

Teams were tested individually. After the arrival of both developers, they were informed about data protection guidelines as every session was completely recorded on video. Following this, the manipulation was introduced. That is, participants read only the instruction, read the instruction and performed cooperative tracking themselves for ca. eight minutes or read the instruction, watched the screen recording and asked questions. In the main part of each session, teams of developers created one or more concepts to assist a cooperative tracking task. To this end, they could use a white board or paper and multi-colored pens to take notes and draw sketches of their ideas. When teams concluded that they had completed creating their concept(s), the experimenter asked them to give a short oral summary of the system developed. This served the purpose of clarifying remaining questions of the researcher. Participants then filled in a short questionnaire on demographic variables. Finally, they received payment and were informed about background and aim of the study. Sessions lasted between one and two hours.

5 Results

5.1 Analysis of Data

Videotapes were transcribed; results of the development processes were inferred from videos, transcriptions, notes taken by participants and notes taken by the experimenter. These results were classified using the taxonomy of assistance systems mentioned before [7]. Some additional categories were inferred from the material using qualitative content analysis [14] and Grounded Theory techniques [15].

5.2 Inter-rater Reliabilities

The majority of the material was classified by one person, the author of this paper. Though, a sample (10 out of 31 teams) of the data was also analyzed by a second independent, trained rater. Inter-rater reliabilities were calculated for a total of 46 variables; Cohen's κ was used as statistical measure. κ ranged from a minimal value of .5 to a maximum value of 1. The mean κ, calculated over all variables, was .851.

5.3 General Results

First analyses were conducted in order to ensure that the three groups did not differ on any of the demographic control variables. An $\alpha = .05$ was set for all following statistical tests. No significant differences between groups could be detected concerning any demographic variable (e. g. age, gender, course of studies, number of projects participated in).

A descriptive analysis revealed the following general pattern: Some action stages seem to be "obvious" ones; these stages are supported by almost every concept and include motivation, activation, and goal setting – warning (30 out of

39 concepts support that stage) and perception (37 out of 39). On the contrary, the stage of information integration/situation awareness seems to be "invisible" as it was supported by no concept at all. Developer teams seemed more inclined to choose an assistance for the stage of decision making (26 out of 39 concepts) than not. Finally, there are stages only few developer teams decide to support. These are action execution (15 out of 39), control of effect (12 out of 39), and motivation, activation, and goal setting – activation and coaching (6 out of 39).

One hypothesis derived from the general hypothesis mentioned above was that teams with more contact with system and users might develop more than one assistance concept. The vast majority of teams created one concept, only few teams created two or three concepts [$M = 1.3,\ SD = .63$]. An analysis of variance did not show any significant difference in number of developed concepts across the three groups [$F(2, 28) = .6,\ p = .58,\ \eta^2 = .007$]. Another effect could be that more contact might result in systems that support a *broader range* of human action stages. Teams in condition two or three might have designed assistance functions that support more stages of human action. An analysis of variance was conducted in pursuing an answer to that question. Yet again, no differences between groups could be detected [$F(2, 36) = .6,\ p = .55,\ \eta^2 = .005$].

5.4 Classification of Data According to Taxonomy of Assistance

Assistance concepts developer teams had created were classified using the previously described taxonomy of assistance systems [7]. Data resulting from these ratings were exclusively nominal. Thus, statistical analyses were based on cross tabulations. As expected frequencies were smaller than 5 in more than 20% of all cases, exact tests were used. Tests were executed for each stage in the human action cycle as well as for every other dimension of the taxonomy. To maximize the amount of information used in statistical analyses, tests were executed for all concepts developed by each team. Thus, this section reports results based on $N = 39$ assistance concepts. As the pattern in results matches those of $N = 31$ concepts—each team is represented by one concept—, the bias introduced by differing numbers per experimental condition was accepted. No significant differences were found between groups concerning any stage of the action cycle, cf. Table 1. Note that no team proposed assistance concepts that could be classified as supporting processes of information integration, i. e. stage three. Thus, no analysis could be executed for that particular stage.

In summary, only one single statistically significant difference between the three experimental conditions occurred concerning the classification of solutions according to a theoretically based taxonomy of assistance. Due to the large number of significance tests conducted, this is probably a random significance. Effect sizes were constantly (very) small, only for two stages (decision making and action execution) the estimated effect sizes were medium. Apart from classifying data according to this taxonomy, other relevant aspects emerging from recorded sessions were classified using schemes derived from data. One of these will be reported to complete the results section.

Table 1. Test statistics (χ^2_{ex}), p-values and effect sizes (ES) for all characteristics of all assistance concepts, $N = 39$

Assistance for Action	Action Stage	Change-ability	Initiative	Medium	Modality	Kind of Input	Parameter
1a) Activation and Coaching	3.1	2.2	2.2	2.0	2.2	3.9	χ^2_{ex}
	.65	1.0	1.0	.94	1.0	.49	p
	.28	.24	.24	.23	.24	.32	ES
1b) Warning	.5	7.6	5.2	12.8	5.2	4.0	χ^2_{ex}
	.90	.17	.22	.51	.22	.10	p
	.11	.44	.36	.57	.36	.32	ES
2 Perception	2.6	7.2	5.8	13.7	3.4	3.4	χ^2_{ex}
	.32	.20	.16	.06	.50	.50	p
	.26	.43	.39	.59	.30	.29	ES
4 Decision Making	13.0	8.1	3.3	16.8	2.6	1.5	χ^2_{ex}
	.06	.15	.52	.047*	.71	.76	p
	.58	.46	.29	.66	.26	.20	ES
5 Action Execution	7.9	10.0	6.2	10.2	6.2	4.0	χ^2_{ex}
	.18	.05	.14	.14	.14	.10	p
	.45	.51	.40	.51	.40	.32	ES
6 Feedback	4.7	5.2	5.2	6.8	5.2	4.0	χ^2_{ex}
	.34	.22	.22	.23	.22	.10	p
	.35	.36	.36	.42	.36	.32	ES

5.5 Assistance for Cooperation

During sessions participants were often observed debating processes of collaboration, cooperation and decision making *between* the two microworld inhabitants. To consider these discussions, the concept of assistance for cooperation was introduced. It was divided into two subdivisions: One aspect was the distribution of tasks, the other aspect was the distribution of decision processes in the team of microworld inhabitants. Each of these two aspects was further subdivided according to the topics discussed by developers.

Distribution of Tasks. Two proposals were by far most popular: (1) Many developers favored an equal distribution of steering between the two microworld inhabitants; this was the default implemented in the system at the time it was presented to developers. (2) An alternative idea frequently generated by developers comprised that one microworld inhabitant should be the pilot and do most of the steering, the other one should be the copilot and act only when necessary. Statistical analyses did not show any significant differences between groups concerning this aspect [$\chi^2_{ex} = 12.9,\ p = .24,\ ES = .6$].

Decision Making in the Microworld. Decision making processes between microworld inhabitants could either not be explicitly arranged. Very few developers opted in favor of letting one of the microworld inhabitants being in command over decisions all the time. Other alternatives included that microworld inhabitants should be informed before a decision situation arises so that they could decide what to do together; the assistance system generates an advice what to do; or microworld inhabitants take turns in deciding. This dimension was the only one in which a significant difference between groups could be detected [$\chi^2_{ex} = 16.5,\ p < .05,\ ES = .7$]. Teams who had fulfilled the cooperative tracking task themselves tended to make no suggestion how to proceed in decision situations, possibly because none of them had experienced difficulties in decision situations. Contrarily, teams in condition three preferred the solution with automatically generated advice on how to decide.

6 Discussion

To sum up the prior results section, with the exception of one variable (cooperative decision support) analyses did not yield any statistically significant differences between the three groups, effect sizes were of small or medium value at best. Thus, *results* of developing processes did not differ significantly according to the manipulation introduced, which was kind of contact with the system.

To interpret these rather devastating results, some remarks should be made. Although there were no significant differences between groups referring to each single stage of action, the general pattern described above was observed. Some action stages seem to be "obvious" ones, other appear to be "invisible". The manipulation introduced did obviously not have enough impact to change that pattern that appears consistently across all groups.

Several other reasons for the lack of expected effects can be given. First, the study reported here was a balancing act between qualitative and quantitative research. Due to restraints of time, only 10, respectively 11, teams could be assigned to each condition. It could be attributed to small size of the sample that in the rare cases of medium effect sizes no statistical significances resulted.

Second, data coded from empirical observations were exclusively categorical data, i. e. measured on a nominal level. This limits possibilities for statistical analyses, only χ^2-tests were reported here. But also more complex analyses using log linear models did not yield different results.

Third, a ceiling effect might have occurred. As the written description of the system and users' tasks in it was very detailed, even teams in the baseline condition who only received that instruction had enough knowledge to provide sufficient solutions. Thus, there was not enough room for the manipulation to result in any observable effects.

Finally, another reason for lacking effects might be adjustment/anchoring effects (cf. [16]). Prior knowledge developers brought into the situation and that was implicitly or explicitly discussed in sessions influenced generation and judgment of solutions. Examples of prior knowledge relevant for this study include current driver assistance systems in vehicles as well as PC driving simulations and racing games like "Need for speed^{TM}". These known solutions constitute an initial value/starting point for developers' judgment of their own assistance concepts. Participants might have been unable to make sufficient adjustments, as such anchors were most likely present.

Generation and judgment processes can also be seen from the point of view of satisficing [17,18]. "Satisficing takes the shortcut of setting an adjustable aspiration level and ending the search for alternatives as soon as one is encountered that exceeds the aspiration level." ([18], p. 13). That is, if a concept created by a developer team reaches the aspiration level—known assistance systems for similar tasks—the team might have concluded its development process.

7 Concluding Remarks and Further Research Plans

Although no major differences between groups could be detected, this study will serve as a basis for two more studies. The 39 resulting assistance concepts will be evaluated in two steps. First, a qualitative expert evaluation of all concepts will be conducted. A small group of experts for creating and evaluating assistance systems will evaluate concepts with regard to how well they support a fast and accurate completion of the task and how they support cooperation and communication between participants. This study will show if the quality of assistance concepts differs and which concepts are judged to be the best ones.

Second, assistance concepts judged as the best ones will be implemented. In a within-subjects design, microworld inhabitants will track cooperatively with assistance given by technical systems and on their own. Time and error rates will be recorded to measure performance. This study will constitute a quantitative evaluation of assistance systems created by developers.

Eventually, performance of teams with assistance given by an operator can be compared to performance with assistance by a technical assistance system created by developers. Thus, a profile of contributions of each of these groups can be assembled. It can be used to show which special division of labor between developer and operator is appropriate for specific situations. If developing and designing activities proceed according to such a scheme of participation, quality of products and software can be improved. Safety for both operators and environment will be enhanced and operators might not have to fight degrading skills any more, but might be able to cooperate with developers and technologies making use of enhanced competences.

Acknowledgments. Research reported herein was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation), as a part of the program Graduiertenkollegs (Research Training Groups). Grant number: Graduiertenkolleg 1013/1 prometei.

References

1. Fitts, P.M.: Human engineering for an effective air navigation and traffic control system. Technical report, Ohio State University Research Foundation, Columbus, OH (1951)
2. Dekker, S.W.A., Woods, D.D.: MABA-MABA or abracadabra? Progress on human-automation co-ordination. Cognition, Technology & Work 4, 240–244 (2002)

3. Sheridan, T.B., Verplank, W.L.: Human and computer control of undersea teleoperators. Man-Machine Systems Laboratory report, Massachusetts Institute of Technology, Cambridge, MA (1978)
4. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part. A: Systems and Humans 30, 286–297 (2000)
5. Kaber, D.B., Endsley, M.R.: The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science 5, 113–153 (2004)
6. Timpe, K.P.: Unterstützungssysteme als interdisziplinäre Herausforderung: Einführung in die Tagung "Wohin führen Unterstützungssysteme?". In: Willumeit, H.P., Kolrep, H. (eds.) Wohin führen Unterstützungssysteme? Entscheidungshilfe und Assistenz in Mensch-Maschine-Systemen, Sinzheim, Pro Universitate, pp. 1–20 (1998)
7. Wandke, H.: Assistance in human-machine interaction: A conceptual framework and a proposal for a taxonomy. Theoretical Issues in Ergonomics Science 6, 129–155 (2005)
8. Billings, C.E.: Aviation automation: the search for a human-centered approach. Lawrence Erlbaum Associates, Mahwah, NJ (1997)
9. Falzon, P.: Enabling safety: Issues in desing and continuous design. In: Design process and human factors integration: Optimising company performance. Proceedings of the Ninth International Symposium of the ISSA Research Section, Nice, INRS (2006)
10. Muller, M.J., Wildman, D.M., White, E.A.: Taxonomy of PD practices: A brief practitioner's guide. Communications of the Association for Computing Machinery 36, 26–28 (1993)
11. Brehmer, B., Dörner, D.: Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. Computers in Human Behavior 9, 171–184 (1993)
12. Gray, W.D.: Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. Cognitive Science Quarterly 2, 205–227 (2002)
13. Sauer, J., Wastell, D.G., Hockey, G.R.J.: A conceptual framework for designinig microworlds for complex work domains: A case study of the cabin air management system. Computers in Human Behavior 16, 45–58 (2000)
14. Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken, 4th rev. edn. Deutscher Studien Verlag, Weinheim (1993)
15. Strauss, A.L., Corbin, J.: Grounded Theory: Grundlagen qualitativer Sozialforschung. Beltz, Weinheim (1996)
16. Tversky, A., Kahneman, D.: Judgment under uncertainty: Heuristics and biases. Science 185, 1124–1131 (1974)
17. Simon, H.A.: Rational choice and the structure of environments. Psychological Review 63, 129–138 (1956)
18. Gigerenzer, G., Todd, P.M., ABC Research Group: Simple heuristics that make us smart. Oxford University Press, Oxford (1999)