Designing for Usability of an Adaptive Time Management Assistant

Julie S. Weber and Neil Yorke-Smith

■ *This case study article describes the iterative design process of an AI-based mixed-initiative calendaring tool with embedded artificial intelligence. We establish the specific types of assistance in which the target user population expressed interest, and we highlight our findings regarding the scheduling practices and the reminding preferences of these users. These findings motivated the redesign and enhancement of our intelligent system. Lessons learned from the study—namely, that AI systems must be usable to gain widespread adoption and retention and that simple problems that perhaps do not necessitate complex AI-based solutions should not go unattended merely because of their inherent simplicity—conclude the article, along with a discussion of the importance of the iterative design process for any user adaptive system.*

*T*his case study concerns our experience designing a system that assists in the process of time management. We are working within the infrastructure of a general-purpose, computerized office assistant named CALO (Myers et al. 2007). We work with a component of CALO called PTIME (figure 1), an intelligent, adaptive scheduling assistant that manages a CALO user's calendar and scheduling requests in a mixed-initiative manner (Berry et al. 2006, 2009).

Our research contributions include (1) highlights from a case study in iterative design in the domain of time management, (2) results of two user studies that motivated the various stages of development of the system, and (3) design implications of these studies, both on the system itself and more broadly.

Efforts to build semiautomated calendaring tools are plentiful (Kozierok and Maes 1993, Modi et al. 2004, Tullio et al. 2002). Among other studies of user scheduling habits and calendaring tools, Palen (1999), notably, examined the use in situ of group calendaring software at Sun Microsystems. Studies report reluctance to invest in accurately and fully informing a scheduling system of preferences (to the extent that an individual can articulate them) unless either (1) the process is not burdensome and people are persuaded of the benefit, or (2) they are mandated to do so (Tullio et al. 2002). It was thus an important prerequisite for the PTIME system to be asto behave in a nonintrusive as possiblemanner, learning user preferences behind the scenes whenever possible. Other studies emphasize that, even when confident in the behavior of a (semi-)automated system, people seek transparency into its reasoning (Glass, McGuinness, and
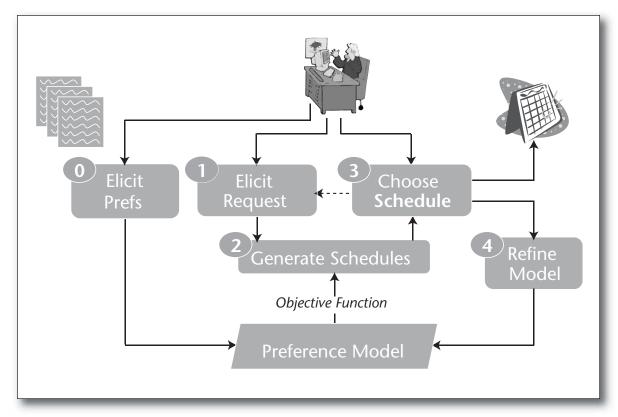
Figure 1: Scheduling Process in PTIME.

The system employs AI reasoning and learning techniques to present likely preferred scheduling options for a given meeting request (Berry et al. 2006).

Wolverton 2004; Palen 1999). PTIME was thus tasked with providing the most appropriate scheduling solutions possiblewhat it found to be the most appropriate scheduling solutions in response to each user query, with multiple efforts, as we will describe, required to achieve success.

We begin with a description of and findings from an initial user study that was conducted to learn about the target users of the CALO system. We then highlight the aspects of system redesign that were prompted by the results of this first study. A second study, conducted at the time of the updated system and described afterward, expanded upon the first study in a manner more focused on obtaining quantitative information regarding user preferences and actual system use. Primary findings are presented in line with each study description, followed by a description of the system design that arose in response to both studies. We conclude by highlighting the lessons learned from our system designs and study results, along with a brief summary and our conclusions.

First User Study

Good practice in designing for usability, for both adaptive and nonadaptive systems, begins with understanding the user population and the nature of the work expected to be accomplished (Gould and Lewis 1985). Due to external factors, the first design of PTIME was driven by technology. Upon our entry to the project, we conducted a mixed-method study to understand the calendaring practices of our target user group of busy knowledge workers. The specific insights we sought to gain through this effort were (1) *what do you do and what is difficult about it?* and (2) *what would you like to do, and where might you appreciate assistance?*

The study consisted of two parts: an in situ diarying by the subjects, and a post hoc evaluation of the diary combined with a semistructured interview.[1] We chose to use in situ dairying because we expected the overhead of diarying and its disturbance to the user's scheduling practices to be minimal, and because it provided a focal point for the subsequent one-on-one interview. We were not seeking substantive quantitative data from the diaries; rather, our aim in the study was to understand the qualitative factors influencing users' scheduling practices.

Our study had included 11 participants, with various office- and research-based roles. We asked participants to keep maintain a diary over one week's scheduling activities, keeping track of how

meetings were scheduled, what decisions were made, and why. We specifically targeted some sub~~jects~~ individuals with crowded, constrained sched~~ules~~. Two of the ~~subjects~~ participants were unable to perform the diarying, either due to excessive busyness (despite the low overhead of the exer~~cise~~), or because their schedules are managed by their administrators. We nonetheless interviewed these ~~subjects~~ participants regarding their schedul~~ing~~ practices. In one case, we asked an administra~~tor~~ to perform the diarying and interviewed her with respect to scheduling for the individual she assists.

The fivefold foci of the study, along with a syn~~opsis~~ of our findings, are as follows:

Event characteristics: People perceive three dimen~~sions~~ to the meetings on their calendars: (1) one-on-one versus group; (2) mandatory versus optional attendance; and (3) fixed versus floating.

Scheduling processes: People schedule events in a variety of ways, including (1) walk-in; (2) constraint satisfaction (group meeting participants' prefer~~ences~~ are elicited until a mutually acceptable meet~~ing~~ time is found); and (3) iterated refinement (a back-and-forth process of finding a mutually acceptable meeting time).

Decision factors: (1) Meeting importance; (2) urgency or criticality; (3) interest; (4) relationships (for example, to host, other participants [Tullio et al. 2002]); (5) perturbation (effect on other meet~~ings~~); (6) stability (that is, the number of times the meeting has been rescheduled); and (7) preferences (of the individual and of others).

Preferences: There were four categories of schedule preferences: (1) meeting-specific (for example, time of day); (2) calendarwide (for example, fragmenta~~tion~~, number of events per day); (3) related to relax~~ation~~ of meeting request constraints (for example, proximity to specified time); and (4) related to relaxing calendarwide constraints (for example, no overlaps).

Scheduling needs: Common suggestions for creating a system to meet its users' needs included (1) coor~~dinating~~ meetings between a group of busy people; (2) intelligent reminders; (3) transparency into assisted scheduling decisions; and (4) greater con~~trol~~ over scheduling processes (for example, accept~~ing~~ requests or not).

Summarizing our findings and those of earlier studies, we divide scheduling tasks into three broad areas: (1) calendar input, (2) calendar view~~ing~~, and (3) event attendance.

In the calendar input area (1)~~task~~, we distinguish (a) the easy case of simply adding an entry to the calendar from (b) the hard case of scheduling or negotiating a meeting or event. In case 1a, the user has the precise entry in mind, and simply adds it, whereas in case 1b the precise entry may be yet unformulated ("I'd like it on Monday"), or may be constrained such that it cannot be entered ideally

(for example, "Bob cannot meet on Tuesday").

The calendar viewing area (2)~~task~~ includes examination and management of one or more cal~~endars~~. Here the user seeks to understand her schedule and allocate her time.

The event attendance area (3~~task~~) involves act~~ing~~ based on calendar entries. Here a user employs the calendar and related artifacts to ensure she acts to meet her objectives, for example, departing on time to attend a meeting.

First Design Iteration

The first iteration of our design focused on devel~~oping~~ capabilities of an assistive agent for the task of arranging a new meeting. Specifically, the design goal was for the hard case 1b of the calendar input task, in which the user cannot simply add the meeting to the calendar, because (1) she does not have the specification of the desired meeting already precisely formulated in her mind (for example, "I want it sometime on Monday"), (2) the specified meeting cannot be coherently arranged as formulated (for example, the specified room is already in use at the specified time), or (3) negotiation with other parties is required ("Bob is not available to meet with you next Tuesday," "Chris prefers afternoon meetings").

At the same time, an important goal was to mesh with the user's preferred calendaring tools and workflow (especially because users have been found to exhibit strong preferences for customiza~~tion~~ and resistance to mandated tools [Glass, McGuinness, and Wolverton 2004; Palen 1999; Tullio et al. 2002]).[2] We did not want to make dif~~ficult~~ the tasks already adequately facilitated by existing tools, namely the easy case (a) of calendar input and calendar viewing. In our first design we did not attempt to assist with event attendance; in particular, we did not address providing reminders.

There were additional requirements on the func~~tionality~~, scope, and user interface of the assistant system. First, it must be a consistent part of the larger overall CALO assistant. Second, it must inte~~grate~~ with existing enterprise calendaring compo~~nents~~. Third, it must include significant automat~~ed~~ learning capabilities. We summarize key elements of the initial design. A fuller description and the rationale of the system at conceptual, functional, informational, and user experience lev~~els~~ is given in Berry et al. (2006).

Manifesting the impact of machine learning is a familiar challenge for self-adaptive systems (Berry et al. 2009; Kozierok and Maes 1993). To help over~~come~~ the "cold start" problem, the PTIME system allows users to describe initial preferences through a "wizard" interface (step 0 in figure 1).

Figure 2 shows one part of a common interac~~tion~~ with PTIME. In the top section of the window
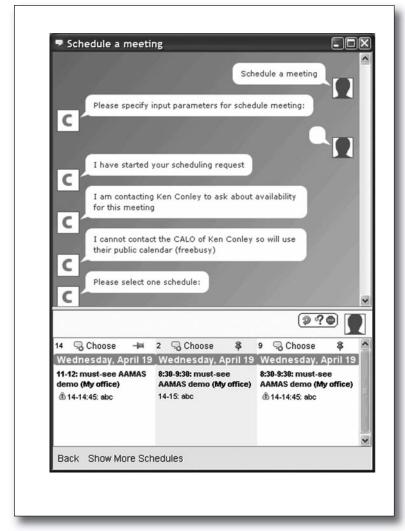
Figure 2. Schedule Option Presentation in the Initial Design of PTIME.

Options are presented in the lower half of the screen as part of the user-system dialogue displayed in the upper half.

we see a system-user dialogue, following DiamondHelp (Rich et al. 2005). In the bottom section of the window we see dynamic, context-specific content, here the presentation of meeting options for the requested meeting.

Leveraging this dialogue metaphor, PTIME operates as an assistant, invokable by the user from within her calendaring application (for example, a button within Microsoft Outlook) or through several CALO-wide interfaces (described in Myers et al. [2007]). PTIME accepts user advice that regulates its autonomy. When tasked by the user with a scheduling request, the system generates and presents a set of options; it does not automatically add the proposed meeting to any calendar unless the request can be met with no relaxations and the system has been granted the autonomy to go ahead. Likewise, participants invited to the meet-

ing must accept the invitation; it is not automatically added.

Meeting requests are given to PTIME through a hybrid natural language form interface (corresponding to step 1 in figure 1; not shown in figure 2). A motivation for employing natural language is to allow the user to specify rich, soft constraints, while not giving her the false impression that she must fill out every aspect of a meeting request. The scheduling domain scopes the statements the user wishes to input so that restricted natural language is practicable, and directs the statements, so that it is intuitive. The interface summarizes the information the user has entered already and stimulates her with "example" lines indicating possible information she could enter. At any time, the user can switch between the hybrid natural language view and a classical form-based view.

Second User Study

The purpose of our second user study was to evaluate the then current system and provide further insights into our target users' general reminding preferences and scheduling practices.[3] This study was undertaken with a greater number of subjects (31 compared to 11), seeking quantitative data rather than qualitative insights and developing breadth in questions over scheduling and reminding preferences rather than depth about scheduling practices alone. The questions we developed were in part motivated by the latent intention to develop a reminding component within PTIME, as well as the desire to improve on earlier designs of the system, that is, further target scheduling task 1b (handling difficult scheduling problems) as well as task 3 (facilitating event attendance).

Scheduling

Our study had a strong focus on understanding participants' scheduling practices and preferences. Questions varied from asking about general calendar management practices (for example, which types of calendars—electronic, paper, mental, and so on—are preferred; how many meetings per week tend to be listed in the calendar) to preferred methods for providing feedback to an automated calendar assistant (for example, willingness to rank a set of toy or true scheduling options; willingness to provide the system with an overall satisfaction rating). Results reinforced the understanding of the space of scheduling practices garnered in our initial user study. This was especially useful due to the larger number of participants; hence, we can be more comfortable generalizing the earlier results on scheduling practices.

Reminding

To address scheduling task 3, namely the facilita-

tion of event attendance, we began to develop an adaptive reminding component of PTIME. This component, not included in the first iteration of the design, provides reminders tailored to user preferences in accordance with a learned model of the user.

Our study, conducted prior to developing the reminding component, and described in detail in Tullio et al.Weber and Pollack (20022008), focused on understanding how people prefer to be reminded and whether there are detectable patterns among user preferences that may be useful for both targeting and expediting the preference learning process. We measured and compared preferences both with and in the absence of explicit contextual information. We presented participants with a set of screenshots representing eight different visual reminders (see figure 3), most appearing on a computer screen and one displayed on a mobile phone. We found that, in the absence of explicit contextual information, preferences for visual reminders were quite similar between subjects: the same types of visual reminders are considered more or less annoying, or intrusive, when no other information is provided. When subjects were asked to consider context, preferences became more diverse: for example, some preferred more intrusive reminders in some instances and not others, and others preferred a certain type of reminder much more exclusively.

We mined the data to gain insights about the relationship between individuals' reminding preferences and scheduling practices. Some interesting correlations exhibited in our dataset include that only a small percentage of people who rely on paper calendars tend to use any type of reminding tools, and that people who use a mobile phone to manage their schedules (perhaps in addition to other means) tend to be most enthusiastic about the prospect of reminder systems to aid their time management.

Second Design Iteration

The second iteration of our design focused on improving the system in the hard case 1b of calendar input, supporting event attendance by providing adaptive reminders, and continuing to keep simple other calendaring tasks. Figure 3 4 shows the options selection interface, which exemplifies the evolution in the design, based on the second user study. While the conversational metaphor remains unchanged, we highlight two changes in the design (Gould and Lewis Berry et al. 19852009).

First, options are presented as a list of "headlines" summarizing the new meeting, approximately 10 to a page, in contrast to the previous side-by-side comparison of three schedules in

Figure 3. Two of the Eight Visual Reminders.

One reminder is displayed on the desktop sidebar and another on a mobile phone.

detail. This display allows the user to see more scheduling options at a glance, and, for each option, to see the proposed new meeting. It is still important to convey the changes to existing meetings—each option is a schedule, not just a new meeting—which was formerly depicted in the details provided. A relaxation of the meeting request, or a modification to existing calendar entries, is indicated by a warning icon; on mouseover, the user is presented with the details. In conjunction, on mouseover of any option, the position of the new meeting is highlighted in a view of the user's calendar (right-most window in figure 34); also shown in this display are details of the preferences of participants and details of relaxation or modifications.

Second, an explicit means to give feedback about any option is available in the form of a rating from one (poor) to four (good) stars. Subjects in the second user study indicated favor to rating options with a simple and familiar mechanism of this type, in contrast to other means of providing feedback. Ratings are optional for the user; on each occasion that an option receives a rating, the information is passed to the PTIME learner to refine its preference model. Implicit feedback in terms of the chosen schedule option is also still obtained.
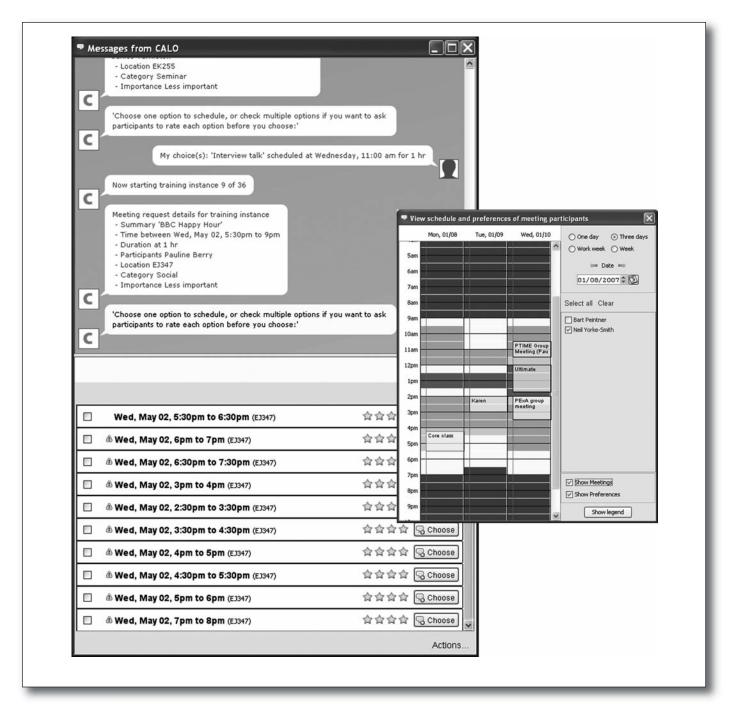
Figure 4. Schedule Option Presentation in a Later Iteration of PTIME Design.

Star ratings permit (optional) user feedback; options are presented in context of a calendar display (right).

Lessons Learned

We highlight two important lessons underscored from our experiences combining AI and human-computer interaction (HCI) as described in this case study. The first lesson is that an AI system must be usable, in the context of the functionality that it provides, in order both to gain users and to gain data to evaluate the merits of the AI. Under-standing and designing for user needs, as described by Spaulding and Weber (2009) in this issue, requires more than just a simple questionnaire or interview. The initial design of PTIME was not conducive to usability. As the design evolved, however, the number of users increased and the quality of the data logged by the system (used to evaluate learning) improved. However, some iterations of our design process could have been condensed

through the use of additional needs assessment techniques and design practices (such as exploratory and participatory design, which incorporates direct user feedback into a variety of system designs). Although seemingly a basic lesson, the importance of usability is too often depreciated in system building compared to algorithmic development.[4]

The second lesson is that usability is more than just the user interface: also needed is the functionality that is useful in the common case, even if that case is less interesting from the AI perspective. Although there was little AI and little novel UI design in parts of PTIME—such as the simple cases of calendaring functionality and nonadaptive reminders—this functionality was essential for ensuring and maintaining usability of the system as a whole.

Summary and Conclusions

In our PTIME system, user input helped us to appreciate that users were having difficulty adding certain events to their calendars and understanding the system's automated scheduling decisions. Iterations on the system design eased these shortcomings and allowed for more seamless time management overall. Rendering an AI system usable requires more than simply creating an interface between the system's back and front ends. Learning from both user behavior (use of the system) and user input (into the design process) led to improved and wider adoption of this intelligent interactive system.

Acknowledgments

Notes

1. As the authors note, diary studies are a relatively inexpensive and low-impact method of capturing user data over a long time period. As is discussed in the usability engineering theme article in this issue, needs derived from the diary data should be confirmed with the end users as was done here with the follow-up interviews.

2. This is an example of the strategies, discussed in the usability side-effects theme article in this issue, of giving users control and limiting learning demands: they can continue to use their preferred tools, rather than be forced to learn a new system.

3. As discussed in the understanding user needs section of the usability engineering theme article, user research should not be thought of as the initial phase of a project but as an effort that continues throughout the project. Insight may be gained every time ideas or prototypes are exposed to users.

4. This is the binocular view that the successful design of interactive intelligent systems should combine both intelligent algorithms and user interaction principles.

References

Berry, P.; Conley, K.; Gervasio, M.; Peintner, B.; Uribe, T.; and Yorke-Smith, N. 2006. Deploying a Personalized Time Management Agent. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, 1564–1571. New York: Association for Computing Machinery.

Berry, P.; Donneau-Golencer, T.; Duong, K.; Gervasio, M.; Peintner, B.; and Yorke-Smith, N. 2009. Evaluating User-Adaptive Systems: Lessons from Experiences with a Personalized Meeting Scheduling Assistant. In *Proceedings of the Twenty-first Conference on Innovative Applications of Artificial Intelligence,* Pasadena, CA, July. Menlo Park, CA: AAAI Press.

Glass, A.; McGuinness, D. L.; and Wolverton, M. 2004. Toward Establishing Trust in Adaptive Agents. In *Proceedings of the Thirteenth International Conference on Intelligent User Interfaces,* 227–236. New York: Association for Computing Machinery.

Gould, J. D., and Lewis, C. 1985. Designing for Usability: Key Principles and What Designers Think. *Communications of the ACM* 28(3): 300–311.

Kozierok, R., and Maes, P. 1993. A Learning Interface Agent for Scheduling Meetings. In *Proceedings of the First International Conference on Intelligent User Interfaces*, 81–88. New York: Association for Computing Machinery.

Modi, P. J.; Veloso, M. M.; Smith, S. F.; and Oh, J. 2004. CMRadar: A Personal Assistant Agent for Calendar Management. In *Proceedings of the Sixth International Bi-Conference Workshop on Agent Oriented Information Systems*, 169–181. Berlin: Springer.

Myers, K.; Berry, P. M.; Blythe, J.; Conley, K.; Gervasio, M.; McGuinness, D.; Morley, D.; Pfeffer, A.; Pollack, M.; and Tambe, M. 2007. An Intelligent Personal Assistant for Task and Time Management. *AI Magazine* 28(2): 47–61.

Palen, L. 1999. Social, Individual, and Technological Issues for Groupware Calendar Systems. In *Proceedings of the Seventeenth Conference on Human Factors in Computing Systems*, 17–24. Pittsburgh, PA. New York: Association for Computing Machinery.

Rich, C.; Sidner, C.; Lesh, N.; Garland, A.; Booth, S.; and Chimani, M. 2005. DiamondHelp: A Collaborative Interface Framework for Networked Home Appliances. In *Proceedings of the Twenty-fifth IEEE International Conference on Distributed Computing Systems Workshops*, 514–519. Piscataway, NJ: Institute of Electrical and Electronics Engineerings.

Spaulding, A., and Weber, J. S. 2009. Usability Engineering Methods for Interactive Intelligent Systems. *AI Magazine* 30(4).

Tullio, J.; Goecks, J.; Mynatt, E.; and Nguyen, D. 2002. Augmenting Shared Personal Calendars. In *Proceedings of the Fifteenth Annual Symposium on User Interface Software and Technology*, 11–20. New York: Association for Computing Machinery.

Weber, J. S., and Pollack, M. E. 2008. Evaluating User Preferences for Adaptive Reminding. In *Extended Abstracts of the Twenty-sixth Conference on Human Factors in Computing Systems*, 2949–2954. New York: Association for Computing Machinery.

Julie Sage Weber is a doctoral student in computer science and engineering at the University of Michigan working in the areas of AI and HCI. Her research interests are in applications of artificial intelligence that benefit society, particularly through assistive technologies and information and communication technologies for development.

Neil Yorke-Smith is a computer scientist at SRI International's AI Center and an assistant professor at the the American University of Beirut. His research aims to help people make decisions in complex situations, with interests that include planning and scheduling, preferences, advisable agents, intelligent interfaces, and constraint programming, and their real-world applications. He holds a Ph.D. from Imperial College London.