

Mixed-Initiative Interface Personalization as a Case Study in Usable AI

Andrea Bunt, Cristina Conati, and Joanna McGrenere

■ *Interface personalization aims to streamline the process of working in a feature-rich application by providing the user with an adapted interface tailored specifically to his or her needs. The mixed-initiative customization assistance (MICA) system explores a middle ground between two opposing approaches to personalization: (1) an adaptable approach, where personalization is fully user controlled, and (2) and adaptive approach, where personalization is fully system controlled. We overview MICA's strategy for providing user-adaptive recommendations to help users decide how to personalize their interfaces. In doing so, we focus primarily on how MICA handles threats to usability that are often found in adaptive interfaces including obtrusiveness and lack of understandability and control. We also describe how we evaluated MICA and highlight results from these evaluations.*

Software applications are often rich in functionality to accommodate users with diverse needs. Consequently, their graphical user interfaces (GUIs) can be complex, more so than is necessary from an individual user's perspective. One means of helping users cope with this complexity is to provide them with a GUI that is personalized to their specific needs (for example, McGrenere, Baecker, and Booth [2002]). In this article, we highlight issues from the design and evaluation of a specific approach to personalization: a mixed-initiative solution, where both the user and system participate in the personalization process. This approach is an instance of using AI to enhance usability by having technology adapt to the user (compare the theme article on usability benefits of AI by Lieberman in this issue), though the adaptation is done with the active participation of the user.

Mixed-initiative strategies (Horvitz 1999) for GUI personalization combine aspects of: (1) adaptive approaches, which rely on AI techniques to personalize the GUI automatically (for example Gajos and Weld 2004); and (2) adaptable approaches, which rely on users to personalize on their own through direct manipulation interface mechanisms (for example, McGrenere, Baecker, and Booth [2002]). In combining elements of adaptive and adaptable approaches, the goal of a mixed-initiative solution is to leverage each of their respective advantages while ideally minimizing their disadvantages. For example, an adaptable interface maintains a very high degree of user control; however, prior research has found that not all users are willing to invest the necessary effort (Mackay 1991) and not all users create the most efficient personalized interfaces (Bunt, Conati, and McGrenere 2004). Purely adaptive interfaces can save the user time but can suffer from a number of usability side effects. A thorough discussion of these side effects can be found in Jameson's (2009) theme article on handling threats to usability (in this issue).

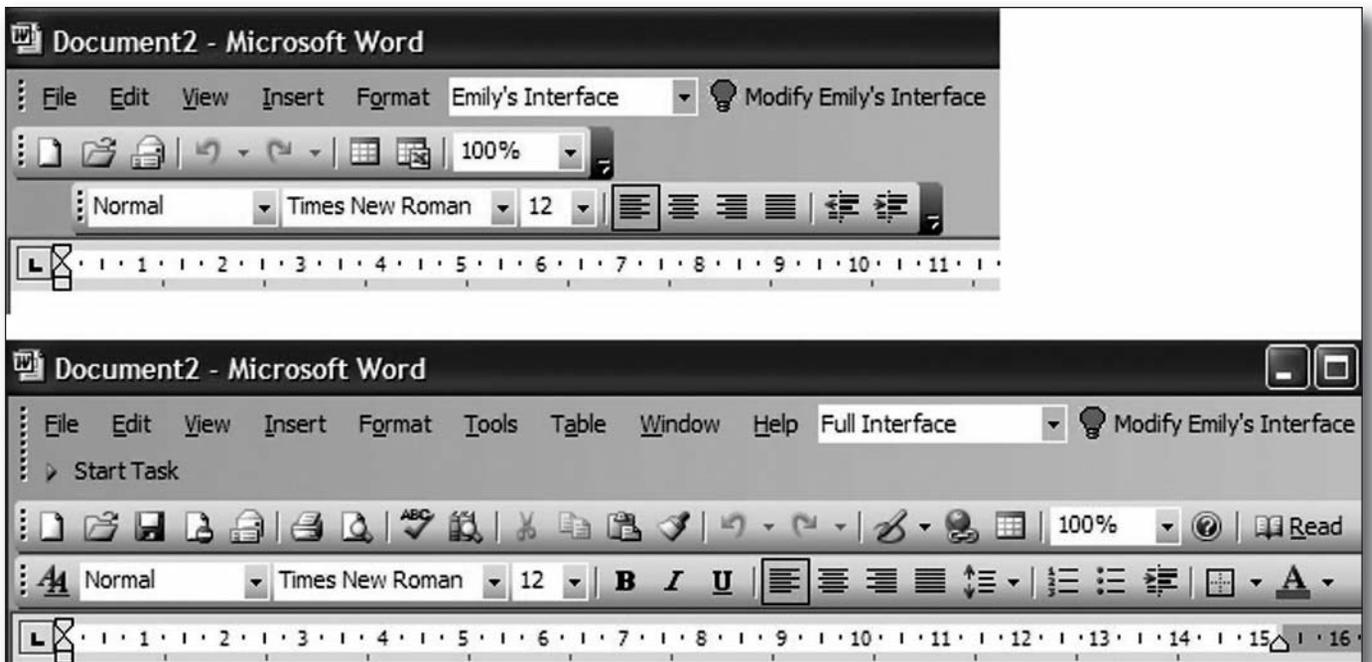


Figure 1. The Two-Interface Model Implemented for Microsoft Word.

The top shows the personal interface. From this interface the users can toggle (using the drop-down menu labeled “Emily’s Interface”) to the full interface, shown at the bottom. Users can add and remove features from the personal interface by clicking the “Modify Emily’s Interface” button.

In this article, we present our mixed-initiative approach to GUI personalization, which is implemented within the mixed-initiative customization assistance (MICA) system. Since detailed descriptions of MICA and its two evaluations can be found elsewhere (Bunt, Conati, and McGrenere 2007a; Bunt, McGrenere, and Conati 2007b), here we focus on aspects of MICA’s design and evaluation that relate to the development of usable AI: We discuss ways in which MICA’s personalization interface aims to deal with three common threats to usability: intrusiveness, lack of control, and lack of understandability. We also discuss key aspects of our evaluation methodology and highlight results from these evaluations.

Overview of the MICA System

MICA provides users with a direct manipulation facility that allows them to decide when and how to personalize, but makes recommendations to help users personalize effectively. Before describing how MICA generates these recommendations, we first describe the type of personalization supported.

MICA supports personalization of the two-interface model proposed by McGrenere, Baecker, and Booth (2002). As shown in figure 1, the two-interface model allows users to work in a feature-reduced personal interface (figure 1, top), but also to access the complete full interface (figure 1, bot-

tom), which they can return to at any point with a single button click. Along with the two-interface model, McGrenere and colleagues designed a lightweight direct-manipulation mechanism that allows a user to add and delete features from his or her interface. Both the two-interface model and personalization mechanism were validated through a six-week field study.

We augment McGrenere and colleagues’ direct manipulation personalization facility with system-generated user-tailored recommendations designed to increase an individual’s personalization effectiveness and decrease his or her personalization effort. A unique aspect of our approach is that MICA’s personalization recommendations are dynamically generated based on a formal assessment of how different interfaces will affect the individual user’s task time. Using a form of cognitive modeling known as GOMS (goals, operators, methods, and selection rules) analysis (Card, Newell, and Moran 1983), MICA simulates and quantifies user performance with different potential interfaces. MICA compares the results of these simulations and makes recommendations based on the interface determined to be optimal in terms of time on task for the particular user in question.

MICA’s reasoning mechanism is designed to factor both interface- and user-specific information into its quantitative GOMS-based simulations. Interface-specific information includes the size and location of all interface features, which affect a

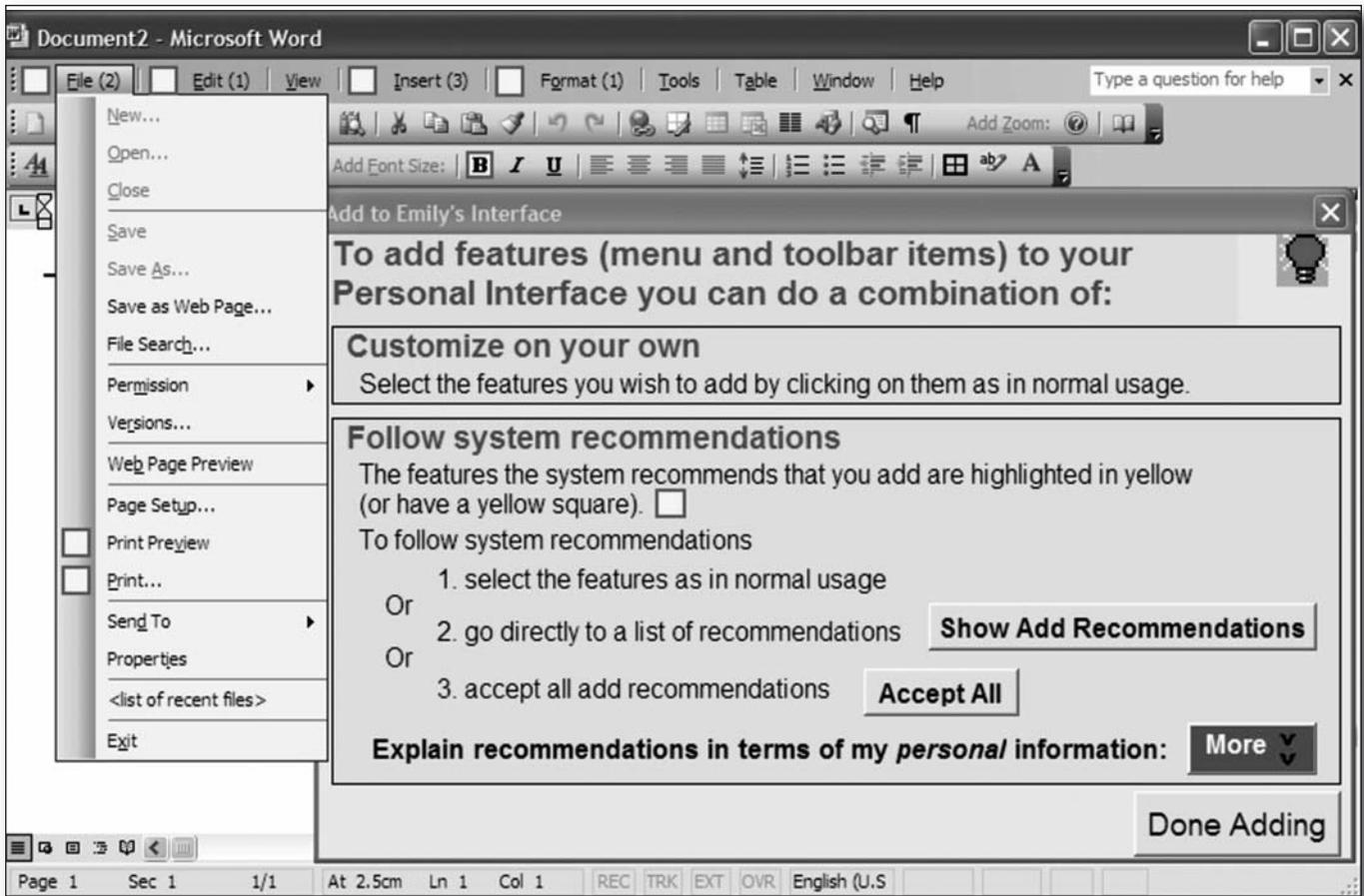


Figure 2. MICA's Mixed-Initiative Personalization Interface.

MICA's recommendations are made visually distinct within the menus and toolbars by highlighting them with yellow squares (seen as gray boxes in the grayscale image).

user's target acquisition time. The user-specific information that MICA can factor into its assessment includes anticipated feature usage and user expertise. The latter factor affects the time it takes users to locate features in the interface given the interface's complexity (for example, Cockburn, Gutwin, and Greenberg [2007]). MICA also factors in an estimate of the cognitive overhead associated with having to switch to the full interface to use any features that are not present in the personal interface. All user-specific information is stored within MICA's user model component.

The MICA system has currently been implemented in the context of Microsoft Word 2003. Its framework, however, is designed to generalize to any GUI, particularly one composed of menus and toolbars.

MICA's Personalization Interface

In this section we describe the design of MICA's personalization interface, focusing on how MICA handles three potential usability threats identi-

fied in Jameson's theme article on usability side effects of intelligent processing in this special issue. Specifically, we describe MICA strategies for dealing with obtrusiveness (called "distraction" in the theme article), lack of control, and lack of understanding.

Handling Obtrusiveness

When working in an application such as Microsoft Word, achieving an effective personalized interface is not the user's primary task. As a result, MICA's goal in presenting its recommendations is to be as unobtrusive as possible. To minimize intrusion, MICA makes recommendations only when the user indicates that she or he is ready to personalize (by clicking the "Modify Emily's Interface" button in figure 1). At this point MICA computes its recommendations using the GOMS-based simulations described in the previous section and displays the dialogue box shown in figure 2.

While this strategy was perceived as unobtrusive in our evaluations, we found that it does not provide sufficient support for those who are interest-

ed in personalizing but are forgetting to do so. It is likely that more proactive system behavior is needed to assist these individuals. Since more proactive support might mean interrupting the user's primary task, great care will have to be taken to ensure that the interruptions come at a good time within the user's task flow and that they are presented tactfully. Planned future work in this area involves extending MICA's framework to reason about the value of personalization versus the cost of interruption and also exploring different designs for the notifications themselves.

Handling Lack of Control

MICA maintains a high degree of user control by leaving the final decision of when and how to customize to the user. Similar to other studies (for example, Jameson and Schwarzkopf [2002]; McGrenere, Baecker, and Booth [2002]); however, during our pilots and two formal evaluations, we observed individual differences in users' desire for control. Therefore, MICA's interface provides users with multiple ways to follow recommendations, each of which differs in its level of control afforded to the user. At one end of the spectrum, users can fully control personalization, with the recommendations simply acting as a visual guide (see the menus and toolbars in figure 2). At the other end of the spectrum, users can delegate most of the responsibility to the system, by following all recommendations with a single button click ("accept all" in figure 2). Between these two ends of the spectrum, users can go directly to a list of recommendations (by clicking the "Show Add Recommendations" button in figure 2) and select from this list.

The disadvantage of having multiple levels of control is the additional complexity that it introduces into MICA's personalization interface. In particular, explaining the three options requires that the dialogue box shown in figure 2 contain a relatively large amount of text. Designing interfaces that permit users to specify their desired level of control without introducing additional interaction complexity is an area for future work.

Handling Lack of Understanding

To handle lack of understanding and, consequently, to increase interaction transparency and predictability, MICA explains the rationale underlying its behavior. Within its rationale component, MICA first explains why it is making recommendations and estimates the amount of time the user would save by having an interface personalized according to the recommendations (figure 3, top). This estimate is made possible by the GOMS simulations used to generate the recommendations. The "why" explanation is followed by a description of the factors that are a part of MICA's reasoning process (figure 3, middle and bottom). Finally,

when relevant, the rationale component provides access to the User Model's assessment for a particular factor (figure 3, bottom).

Similar to users' desire for control, we observed individual differences in users' desire to have the explanation facility present and its impact on users' attitudes. The explanations increased feelings of trust, understanding, and predictability for some users. Other users, however, found the interaction to be sufficiently transparent and predictable without the explanations or didn't perceive transparency and predictability to be important within this setting. Given the importance placed on these properties in the literature (for example, Jameson [2008]), we were surprised to hear the latter sentiments expressed. These users did, however, mention two characteristics of this particular type of interaction that appeared to contribute to their reactions. First, given that interaction is mixed initiative, users can ignore recommendations that they either do not understand or do not feel to be useful. The second characteristic relates to the fact that users are working in a productivity application with the primary goal of getting their work done. Personalization itself is only a secondary task; some users are keen to complete it as quickly as possible and feel that they have little time to read detailed explanations.

To accommodate the above individual preferences with respect to the utility of the explanations, MICA's rationale component is accessible from the main mixed-initiative dialogue box, but not displayed until requested by the user (by pressing the "more" button in figure 2). Despite our use of an iterative design and evaluation process to formulate the rationale component and help ensure its clarity, there is still room to improve its design and, we hope, increase its utility, particularly for those in a hurry. For example, it may be possible to reduce the amount of text in the explanations and perhaps make them more graphical.

Evaluations

Spaulding and Weber's (2009) theme article on usability engineering methods (in this issue) describes challenges associated with evaluating intelligent systems. In this section, we describe how we evaluated MICA and what we learned from these evaluations.

We performed informal, think-aloud style usability testing throughout MICA's design process to ensure its usability. For more formal validation of MICA's approach, we conducted two laboratory evaluations. In both evaluations, participants took part in all experimental conditions. We used this type of design to (1) reduce the number of acquired participants, (2) account for variability owing to individual differences, and (3)

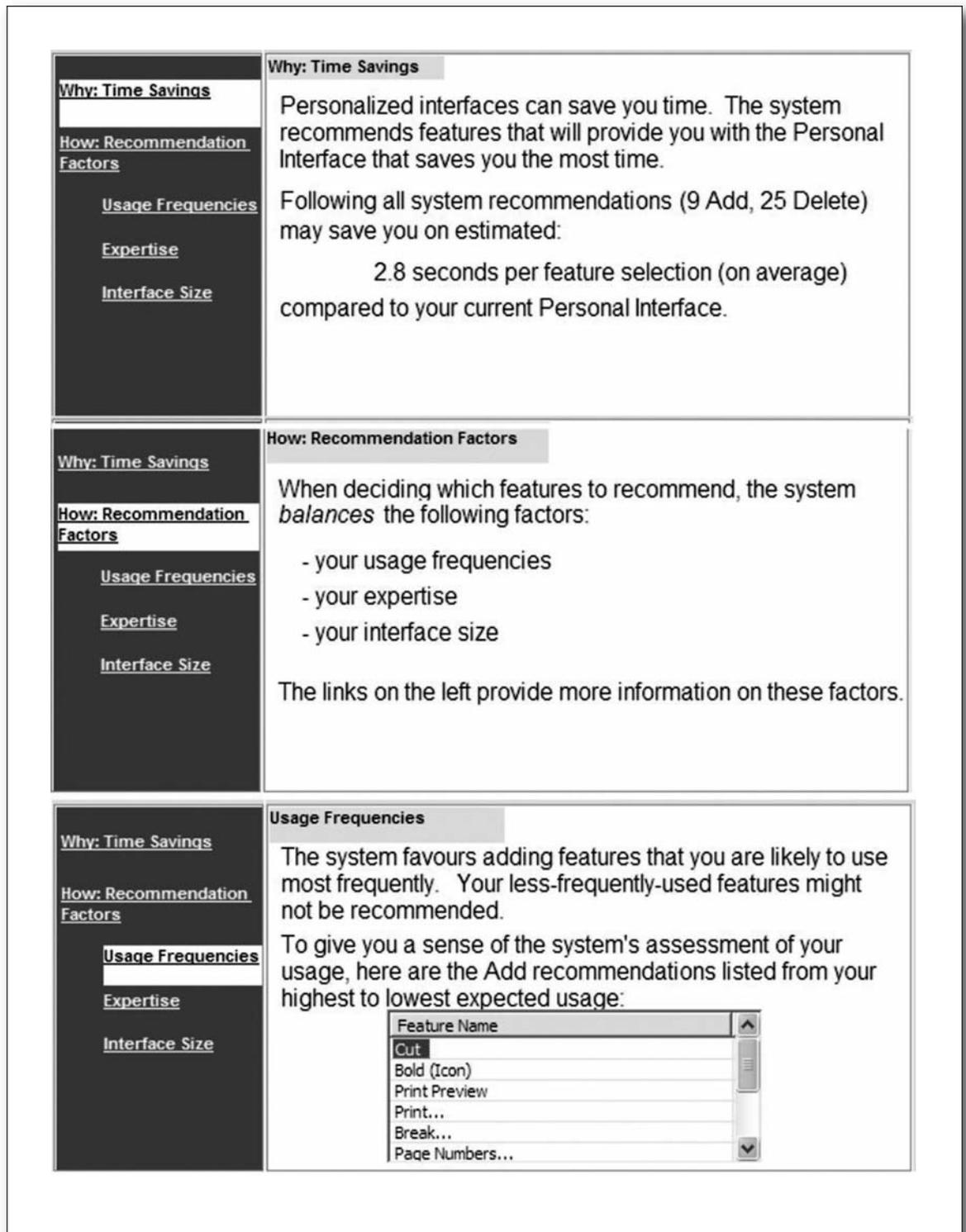


Figure 3. Part of MICA's Rationale Component.

The top screen describes why MICA makes recommendations; the middle screen lists the factors it considers; the bottom describes the "Usage Frequencies" factor and provides access to the User Model's assessments for that factor.

elicit direct comparative statements in regard to the different versions tested.

System Fidelity for User Testing

With research that combines AI and human-computer interaction (HCI) perspectives, a key issue is how much of the system to implement prior to the first set of evaluations. On the one hand, developing a complete system that can capture, assess, and reason about relevant user characteristics has the potential to be an interesting AI contribution. On the other hand, early evaluation of a low-fidelity prototype, as is common in HCI, can generate considerable insight into the advantages and disadvantages of the proposed adaptive behavior, with far less implementation time.

In regards to this tension, we implemented MICA's general framework, which performs comprehensive performance-based reasoning to recommend suitable personalizations; however, we conducted our two evaluations without a fully functional user model. Recall that MICA's user model includes information on user expertise and anticipated feature usage, which MICA uses to make its recommendations. To validate the framework prior to implementing appropriate online assessment algorithms within the user model, we used a variant of the "Wizard of Oz" technique. Participants were given the impression of interacting with a fully functional system; however, user expertise and anticipated feature usage were initialized in the user model by (a) administering detailed expertise questionnaires and (b) having users complete scripted tasks, which gave us fairly accurate information on feature usage.

The danger in evaluating at this stage is that the system might be less usable when working with the actual, likely less accurate, User Model. One solution is to evaluate the system at multiple predetermined levels of accuracy (for example, Tsandilas and Schraefel [2005]), but this can lead to an infeasible number of experimental conditions (particularly for a within-subjects design). If it is not feasible to evaluate at multiple levels of accuracy, ideally one would want to introduce "realistic" noise into the User Model's assessments. Challenges with this approach include determining not only the type and frequency of system errors, but also how to make these errors seem plausible, even in a laboratory setting.

Overview of Findings

Twelve users participated in our first evaluation, which compared MICA to a purely adaptable alternative (Bunt, Conati, and McGrenere 2007a). The purely adaptable version was simply the personalization mechanism shown in figure 2, but without access to any recommendations from the system.

The results indicated MICA had positive impacts on task time and that personalization time with MICA was significantly less than in the adaptable version. Users also preferred MICA to the adaptable version on a number of qualitative measures. In particular, users preferred MICA primarily because it saved them time, but also because it helped them make good personalization decisions.

The second evaluation, which had 16 participants, compared versions of MICA with and without the explanation facility (Bunt, McGrenere, and Conati 2007b). The qualitative results indicated that the majority preferred to have the explanations present for feelings of increased trust, understanding, and predictability. But as discussed earlier, there were also some neutral and negative reactions to the rationale component. Quantitatively, having the rationale present did not significantly affect the number the recommendations followed, with participants following over 90 percent of MICA's recommended additions in both conditions. As one might expect, personalization time did increase as a result of viewing the rationale; however, we do not expect that users would need to view the rationale every time they personalize.

Conclusions

In this article we presented the MICA system—a case study in the design and evaluation of a system that aims to provide users with usable AI to help them personalize their graphical user interfaces. Our evaluations provide encouraging evidence not only for our specific approach but also for mixed-initiative approaches in general. Given the right interaction design, incorporating machine intelligence has the potential to augment the user's experience with few downsides. The nature of our Wizard-of-Oz evaluations means that we tested machine intelligence that was operating at a fairly accurate level; however, we are cautiously optimistic that our results would generalize to a system exhibiting reasonable, even if occasionally error-prone, behavior.

Our case study also highlights the importance of accommodating individual preferences concerning the way in which the system's intelligence manifests itself in its interaction with the user (compare the discussion of controllability in the usability side-effects theme article). In our evaluations, we saw a great deal of variability in both users' desire for control and the degree to which they wanted the system to be understandable. We believe that the ability to cater to these individual differences is a key component of MICA's success. Incorporating such flexibility in an intelligent interactive system, however, is not without cost, which in our case was increased complexity within MICA's mixed-initiative interface.

Future Work

In addition to exploring alternative ways of accommodating individual preferences with respect to desire for control and understandability, there are several other promising avenues of future research. First, we would like to explore alternative designs for MICA's rationale component and more generally ways to design explanations that accurately describe complex adaptive behavior, yet are still lightweight. It is also necessary for the community to gain a better understanding of when and why detailed explanations of adaptive behavior are necessary. Second, we would like to explore alternative mixed-initiative strategies, including one where the system engages in more proactive behavior. Finally, we would like to explore ways of introducing realistic noise into Wizard-of-Oz style evaluations, so that designers of interactive intelligent systems can gather valuable usability data prior to implementing fully functional AI components.

References

- Bunt, A.; Conati, C.; and McGrenere, J. 2007a. Supporting Interface Customization Using a Mixed-Initiative Approach. In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI 2007)*, 92–101. New York: Association for Computing Machinery.
- Bunt, A.; Conati, C.; and McGrenere, J. 2004. What Role Can Adaptive Support Play in an Adaptable System? In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI 2004)*, 117–124. New York: Association for Computing Machinery.
- Bunt, A.; McGrenere, J.; and Conati, C. 2007b. Understanding the Utility of Rationale in a Mixed-Initiative System for GUI Customization. In *Proceedings of the International Conference on User Modeling (UM 2007)*, 147–156. Berlin: Springer.
- Card, S. K.; Newell, A.; and Moran, T. P. 1983. *The Psychology of Human-Computer Interaction*. Mahwah, NJ: Erlbaum.
- Cockburn, A.; Gutwin, C.; and Greenberg, S. 2007. A Predictive Model of Menu Performance. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2007)*, 627–636. New York: Association for Computing Machinery.
- Gajos, K., and Weld, D. S. 2004. SUPPLE: Automatically Generating User Interfaces. In *Proceedings of the ACM Conference on Intelligent User Interfaces (IUI 2004)*, 93–100. New York: Association for Computing Machinery.
- Horvitz, E. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1999)*, 159–166. New York: Association for Computing Machinery.
- Jameson, A. 2009. Understanding and Dealing with Usability Side Effects of Intelligent Processing. *AI Magazine* 30(4).
- Jameson, A. 2008. Adaptive Interfaces and Agents. In *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, 2nd ed., eds. J. Jacko and A. Sears, 433–458. Boca Raton, FL: CRC Press.
- Jameson, A., and Schwarzkopf, E. 2002. Pros and Cons of Controllability: An Empirical Study. In *Proceedings of Adaptive Hypermedia (AH 2002)*, 193–202. Berlin: Springer.
- Lieberman, H. 2009. User Interface Goals, AI Opportunities. *AI Magazine* 30(4).
- Mackay, W. E. 1991. Triggers and Barriers to Customizing Software. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'91)*, 153–160. New York: Association for Computing Machinery.
- McGrenere, J.; Baecker, R. M.; and Booth, K. S. 2002. An Evaluation of a Multiple Interface Design Solution for Bloated Software. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2002)*, 163–170. New York: Association for Computing Machinery.
- Spaulding, A., and Weber, J. 2009. Usability Engineering Methods for Interactive Intelligent Systems. *AI Magazine* 30(4).
- Tsandilas, T., and Schraefel, M. D. 2005. An Empirical Assessment of Adaptation Techniques. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2005), Extended Abstracts CHI, 2009–2012*. New York: Association for Computing Machinery.

Andrea Bunt is an assistant professor in the Department of Computer Science at the University of Manitoba. She obtained her Ph.D. in computer science from the University of British Columbia (2007), after which she completed a postdoctoral fellowship at the University of Waterloo (2007–2009). Her research interests include intelligent interactive systems, personalization, user modeling, and human-computer interaction. More details can be found on her website (www.cs.umanitoba.ca/~bunt).

Cristina Conati is an associate professor of computer science at the University of British Columbia. She received a “Laurea” degree (M.Sc. equivalent) in computer science at the University of Milan, Italy (1988), as well as a M.Sc. (1996) and Ph.D. (1999) in intelligent systems at the University of Pittsburgh. Her areas of research include adaptive interfaces, user modeling, affective computing, and intelligent tutoring systems. She was program cochair of User Modeling 2007, the 11th International Conference on User Modeling, and conference cochair of IUI 2009, the International Conference on Intelligent User Interfaces. She published over 40 strictly refereed articles, and her research has received awards from the International Conference on User Modeling (1997), the International Conference of AI in Education (1999), the International Conference on Intelligent User Interfaces (2007), and the Journal of User Modeling and User Adapted Interaction (2002).

Joanna McGrenere is an associate professor in computer science at the University of British Columbia, where she studies human-computer interaction with a particular focus on user interface personalization, assistive technologies, and groupware technologies. She earned her Ph.D. from the University of Toronto in 2002, and has been an IBM visiting scientist since 2003. More details can be found on her website (people.cs.ubc.ca/~joanna).