User Interaction with User-Adaptive Information Filters

Henriette Cramer[1], Vanessa Evers[1],Maarten van Someren[1], Bob Wielinga[1], Sam Besselink[1], Lloyd Rutledge[2], Natalia Stash[3], and Lora Aroyo[4]

[1] University of Amsterdam, Human Computer Studies Lab, Kruislaan 419, 1089 VA Amsterdam, The Netherlands
hcramer@science.uva.nl
[2] Telematica Instituut, P.O. Box 589, Enschede, The Netherlands
[3] Vrije Universiteit Amsterdam, De Boelelaan 1083a, Amsterdam, The Netherlands
[4] Technische Universiteit Eindhoven, P.O. Box 513, Eindhoven, The Netherlands

Abstract. User-adaptive information filters can be a tool to achieve timely delivery of the right information to the right person, a feat critical in crisis management. This paper explores interaction issues that need to be taken into account when designing a user-adaptive information filter. Two case studies are used to illustrate which factors affect trust and acceptance in user-adaptive filters as a starting point for further research. The first study deals with user interaction with user-adaptive spam filters. The second study explores the user experience of an art recommender system, focusing on transparency. It appears that while participants appreciate filter functionality, they do not accept fully automated filtering. Transparency appears to be a promising way to increase trust and acceptance, but its successful implementation is challenging. Additional observations indicate that careful design of training mechanisms and the interface will be crucial in successful filter implementation.

Keywords: user-adaptive systems, information filtering, transparency, trust, acceptance, recommenders.

1 Introduction

In crisis situations it is vital to have the right information at the right place at the right time. Emergency management personnel, possibly from multiple organisations, need to work together and need to make sense of often dynamic, chaotic and unexpected situations. They have to deal with high-risk situations and both information over- and underload. A diversity of actors needs to be provided with the information they need. Especially in case of international crises, substantial personal and cultural differences might have to be overcome. User-adaptive information filters have been suggested as a possible tool to deliver crisis management actors with the information they need in a personalised, effective and efficient manner (e.g. Meissner, 2006, Van Someren, 2004). However, user interaction with user-adaptive filters is not yet completely understood. The dialogue between a user-adaptive information filter and the user is

N. Aykin (Ed.): Usability and Internationalization, Part II, HCII 2007, LNCS 4560, pp. 324–333, 2007.

extremely important in achieving filtering adequate performance and acceptance in the user (Waern, 2004, Höök, 2000, Hanani, 2001). This dialogue needs to build an appropriate level of trust in the user. Users need to decide how well the system is suited for use in the task at hand and in what situations they can or cannot depend on the system. Over time, users need to assess how well the filter is adapting well to their feedback and is improving its filtering results and possibly becoming more suited to these tasks – or less suited when adaptation is less successful. A relationship needs to be built where the user trusts the system and feels it is useful to invest effort in training the filter, even though the filter might not yield high-quality results from the first moment on. The filter has to convince the user to keep using the filter and provide feedback on its filtering results so the system can improve its future results. This interaction needs to be satisfying to the user, while not detracting from the user's task. Additionally, usability concerns for adaptive systems might have to be addressed, such as limited predictability, controllability and obtrusiveness (Jameson, 2003). Such complexities of user interaction with adaptive systems will in all probability even be greater in crisis situations. Filters will have to adapt to very dynamic, complex situations and the stakes for users are high; achieving appropriate user trust will be crucial. Before user-adaptive information filters are deployed in high-risk situations to address the needs of a diverse set of users we need to understand what factors affect trust in and acceptance of such systems in general. The research discussed in this paper contributes to research in user diversity and usability by aiming to explore the factors that affect trust in automation and the effects of transparency across various types of user contexts. After a short discussion of trust and transparency, two studies are discussed as examples of such research

Parasuraman et al. (2000) propose a model of levels of reliance that could help deciding what level of automation is appropriate in a certain situation to ensure effective performance. They consider human performance areas (mental workload, situation awareness, complacency, skill degradation) and evaluative criteria such as automation reliability, risks and the ease of systems integration. However, even when an appropriate level of automation has been determined, there is no guarantee users will trust a system to the same appropriate degree and will actually decide to use it. Lee and See (2004) provide a comprehensive review of trust from various perspectives (including organisational, sociological, and interpersonal). They define trust as 'the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability'. A similar definition is provided by Jøsang, (2004), who define trust as 'the extent to which one party is willing to depend on somebody or something, in a given situation with a feeling of relative security, even though negative consequences are possible'. An appropriate level of trust should match the reliability of the technology. Muir (1994) however shows that automation is distrusted by users, even if trust would be warranted. Trust is mostly determined by perceived competence of a system. Muir found that any signs of incompetence diminished trust, even when performance was not hindered. However, trust is not just determined by the perceived reliability of a tool. Factors such as perceived risks of using a system, previous experiences with the automation, user's workload, 'good manners' of the automation such as confirming to human etiquette, match of the user's personality and the system's are examples of factors that play a role as well (Parasuraman, 2004, Lee and See, 2004). Furthermore, the information items the

system presents to the user need to be trustworthy as well, focus here is however on trust in the system itself. Lee and See conclude that trust does influence whether users are willing to rely on a system, but that it does not determine reliance completely. They state that trust only plays a role in uncertain and complex situations, where exhaustive evaluation whether to use a system is impractical. Assessing whether a system can be trusted in a complex situation can be made easier by making a system transparent. This entails offering the user insight in how a system works and increasing user understanding of the system's inner workings, for example by offering explanations for filtering results. In previous studies it is suggested that making an adaptive system more transparent to the user could lead to increased trust and acceptance, an increase in system performance and a more positive user attitude towards using a system (e.g. Herlocker, 2000; Pu 2006, Sinha, 2002; Höök, 2000; Waern, 2004; Cortellessa 2005). What types and levels of transparency are feasible and appropriate in which contexts and what type of effects they have on trust and acceptance is not yet fully understood. As Fogg (2003) states, the impact on credibility of any element depends on to what extent it is noticed (prominence) and what value users assign to the element (interpretation). Parasuraman (2004) notes that as automation gets more complex, users will be less wiling and able to learn about the mechanisms that produce the automation's behaviours. Besides the effort needed from the user to process transparency information, additional possible adverse affects have to be considered as well. Dzindolet (2003) for example found that explaining why errors might occur increased trust in automated tools, even when this was not appropriate. Even though transparency is thought to increase trust, there also might be a conceptual limit to the effect on usage behaviour. Trust is only thought to be an issue in situations that are too complex to understand for the user (Lee and See, 2004). If a system is fully transparent, and the user has the resources to do so, s/he can fully understand the system. It could be argued then that in such a situation, trust no longer plays a role.

2 Problem Statement

Understanding what factors affect how users provide feedback to a system and the mechanisms that lead to trust and acceptance of filters in general are important to be able to design user-adaptive information filters suitable for crisis situations. The relationship between risk, trust and transparency is not yet fully understood, especially not in the context of user interaction with user-adaptive systems. From the discussed literature we expect that transparency both influences perceived competence of a system through users' understanding of a system, and possibly also directly influences trust. The research project aims to further explore the factors influencing acceptance/trust in adaptive information filters. How do user perceptions of risk, system transparency and user control affect acceptance and trust? In what situations does transparency affect trust and acceptance? We aim to offer guidance on how a system can offer appropriate transparency and user control and benefit users, their task and context. While the characteristics of crisis situations include a diverse set of multiple users, high-risk and complex situations, we first aim to understand single user situations in less critical contexts and build on our findings in further research.

3 Case Studies

Below a selection of results from two studies are discussed. These studies have been carried out to investigate the relationship between training, transparency, trust and acceptance in interaction with user-adaptive information filters. Both studies address interaction with user-adaptive information filters that are dependent on explicit user feedback and base their filtering on content features of information items. First a study is discussed exploring user interaction with a trainable spam filter. This study investigates which factors affect user attitudes to information filter use in practice. The second study explores user interaction with a content-based user-adaptive art recommender system. This study focuses on the effects of transparency on trust and acceptance. Possible implications of the findings of both studies are discussed for designing an information filter for use in crisis situations. The studies are part of a larger research program investigating interaction with user-adaptive information filters.

3.1 Spam Filter Study

The first example study aimed to investigate what factors in practice influence use, trust and acceptance of adaptive systems that rely on direct user involvement. Interaction with user-adaptive spam filters was chosen as a case study. 48 participants were observed while interacting with their email client and spam filter in their regular work setting. Both participants with and without knowledge of AI-techniques were involved, 30 male, 18 female. Participants were relatively well-educated, ranging from some college-level education to PhD's. All participants were employees of either a university or a research organisation, ranging from researchers to administrative personnel. Use of two types of filters was observed: the built-in Bayesian trainable spam filter of the Mozilla email client, and an additional non-adaptive rule-based filter. After observation, interviewing addressed participants' understanding of their spam filter, and explored their attitude towards using and training their filters. The questionnaire addressed participants' trust and acceptance of their spamfilter, with questions adapted from Venkatesh, (2003) and Jian, (2000).

The study showed that the choice of users for a specific filter setting mainly depended on whether participants thought information overload spam was a problem for them and the way they wanted to manage the risk of losing email. 23 of the 44 participants (out of 48) with an active spam filter spam considered spam a problem, with 14 mentioning that spam wasn't a problem anymore because of their spam filter(s). However, even while the majority of these participants thought their spam filter was useful (average=5.3, st.dev=1.1, N=46, a=.869 on a Likert-scale from 0-6) and had a positive attitude towards the filter (mean = 5.7, st.dev 0.7, N=48), letting the filter actually delete messages was not acceptable to most of them. Only four participants let their filter automatically delete (a portion of) the spam they received. Participants based their level of trust mainly on observed filter errors. Some participants gave examples of 'critical errors' where the filter did not recognize obvious spam or filtered out email that was important to them. Participants reported they trusted their filter in the interview, but scored only slightly positive on trust-related items in the questionnaire (mean= 3.1, st.dev=1.3, N=46, a=.733). The way

participants managed the risk of losing an email appeared decisive in the choice for a particular filter setting. The only participant who let the filter delete all alleged spam for example reasoned that missing an important email was a risk he could easily take, as senders of really important emails would contact him if he didn't answer them anyway. Other participants did not feel this was a viable option for them. Such differences thus occur even within one organisation. Diversity of users and their (social) context and the way these affect the way in which users rely on automation have to be taken into account when designing an information filter. Especially when designing for crisis management, where multiple organisations, a wide diversity of actors and changing task contexts will occur, this will be a major design challenge.

Participants were willing to train their filters and both correcting and more extensive training efforts such as importing old spam as training set were reported. Interestingly, participants appeared to report that training the filter by correcting its errors did appear to have increased trust, but they did not report less controlling behaviour. These user efforts thus mainly appear to have cost the users; they did not benefit from the filter's capabilities. This extra effort spent on a filter instead on the task at hand, without any benefits the user takes advantage of later on, might be fatal in more critical situations.

Fig. 1. a. Mozilla Junk icon b. Mozilla Junk button and c. Mozilla warning text and correction button

Careful interface design is very important in raising awareness of filter activity and avoiding user mistakes in training a filter. Even though the Mozilla client shows icons, buttons and warning labels (figure 1), participants did not always recognize the filter's activity. Such lack of awareness of filter activity might have grave consequences in crisis situations. If filter errors occur, users might do not even realise a breakdown in communication has occurred, or are even possible, which might lead to uninformed actions and inappropriate reliance on automation. One participant in this study for example confused the filter's button to train the filter for a delete button, essentially training the filter to recognize any email similar to any email ever deleted by that participant as spam. This could potentially be very dangerous in more critical situation when a user decides to actively use the filter, without realising it has been wrongly trained. If even in a situation such as in this study, where participants user their email client everyday, they are not always able to report filter settings and sometimes confuse interface elements, this certainly has to be taken in account for crisis situations. Mistakes appear to be even more likely in less-familiar crisis situations in which users will not be able to spend time to fully examine a filtering interface. It appears that while users do appreciate the concept of adaptive filters, and are willing to invest effort to train them, filters are not used to their full extent mainly because of the perceived risks associated with their use and imperfections in interface design leading to lack of awareness and understanding of the filter.

3.2 Transparency, Trust and Acceptance of a User-Adaptive Recommender System

Transparency, giving the user insight in the inner workings of a system, could be a way to increase trust and acceptance of filters when filter usage will actually benefit task performance. The second study described in this paper is a pilot study, part of an ongoing study, exploring how transparency affects user trust in and acceptance of user-adaptive content-based information filters. An art recommender prototype, the CHIP system was used in this study as an example of a user-adaptive information filter. The CHIP system recommends artworks from the collection of the Rijksmuseum Amsterdam based on annotations of content features of artworks. Annotations used for recommendations include for example artist (for example Rembrandt), place & time (such as Amsterdam, 1700-1750), and topics (such as 'food and drink', or Buddhism).

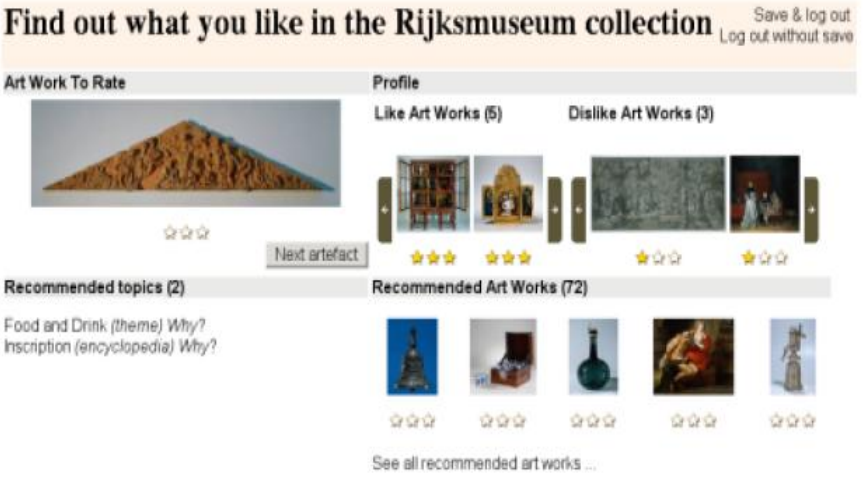

Fig. 2. Transparent version of the CHIP prototype

Using a between-subject experimental set-up the effects were investigated of offering a more transparent versus a less transparent recommender on trust and acceptance. Two groups of participants individually used either a transparent or a non-transparent version of the CHIP recommender system. The transparent CHIP version showed the topics in the user profile the system thinks the user finds interesting and on which its art recommendations are based (figure 2). The non-transparent version did not show these topics. Observation of participants' interaction with the system, interviews and a questionnaire were combined to gain insight in the effects of transparency on perceived understandability, perceived competence, trust and acceptance of the system and its recommendations. Questionnaire items were adapted from Jian (2000) for trust and Venkatesh (2003) for acceptance.

Fifteen participants took part in the study. Participants were 14-61 years old, 10 male, 5 female. Thirteen of the 15 participants had received some form of college

level education. Participants were asked to prepare a presentation about their personal art interests. For this presentation they needed to choose 10 artworks as their favourites. These artworks could be any artwork they had seen using the CHIP system, either recommended or not. If participants chose more recommendations over other artworks they had seen in the system, this was taken as an indicator of acceptance of the system's recommendations. Afterwards, participants were presented with an 'acceptance task'. They were asked to find one additional interesting artwork within one minute and were offered the choice to either find this artwork from a fixed list, not based on their profile, or use the recommended artworks.

Trust, Acceptance and Transparency Effects. Participants generally thought the recommender was moderately useful, scoring a mean of 4.93 on a 7-point Likert scale (st.dev=1.43, N=15; questionnaire items=2, a=.7396). They liked system recommendations (mean=5.23, st.dev=1.18; items=2, a=.6998) and reported trusting the filter (mean=5.2, st.dev=.666, N=15; items:10, a=.869). However, 10 of the 15 participants still reported they would not let the system choose artworks for them. Transparency did not increase trust and acceptance in a significant way in this study. There were also no differences between the conditions in the levels of trust reported in the questionnaire. Participants in the transparent condition did not prefer the system's recommendations to their own selection more often than those in the non-transparent condition. Participants in the transparent condition were even significantly less convinced the system's recommendations were improving using their feedback than participants in the non-transparent condition (U=12.500, W=40.500, Z=-1.83, p(1-tailed)=.036, Ntrans=7, Nnon-trans=8).

A number of possible explanations are possible for the lack of effect of transparency on trust and acceptance, and its unexpected additional negative effects on the perceived benefits of training the system in this study. First of all, the number of participants in this exploratory study was small and the manipulation might not have been strong enough. There was no significant increase in perceived understanding of the system in the transparent condition in this study (U=26.00, W=62.00, Z=-239, p(1-tailed)=.434). This could be explained by the potential unfamiliarity of the topics shown. If the user for example does not understand what a topic such as 'inscriptions' entails and why it would be interesting, showing this topic does not necessarily increased perceived understanding of the system's criteria. Additionally, not all features of the transparent version were noticed by participants. A 'why?' button was offered for every listed recommended topic, clicking this button opened a screen explaining what ratings this recommendation was based on. This feature was not used by any of the participants. Whether they didn't notice or did not have the need for this information is unclear. This illustrates that it is important to make sure that transparency features are actually helping the users understand the system better and not just require attention and mental effort from the user. Participants in the transparent condition did find learning how to use the system more difficult (U=13, Z=-1,93, p=0,05; mean transparent condition=6.00, mean non-transparent condition=6.75). For many participants, the transparency feature became an extra screen item that had to be processed instead of a helpful feature. This type of issue certainly has to be taken into account in designing information filters for crisis situations. A full understanding of the risks and benefits of using a filtering system

might be crucial in deciding whether using the system would be appropriate in a crisis situation. In emergency management applications however users cannot be expected to notice unfamiliar screen features or spend the effort to get to know how a system works during a crisis. It appears that instruction of potential users on how a system works cannot be avoided, if it has to be made sure users can make informed decisions on whether the system is suitable for the tasks at hand. It also appears that there are limits to the effects of transparency even if it makes the system more understandable. Trust and acceptance in this study appeared more determined by the immediate evaluation of the recommendations by participants and not by the type of transparency offered here.

4 Discussion and Conclusion

Participants in the discussed example studies appear to value the concept of user-adaptive information filters. Participants in the first study thought their spam filter helped them deal with a spam problem. Participants in the second study reported to appreciate the recommendations made by the system. However, these studies also show that most users do not want to hand over complete control to an adaptive system. In case of the spam filter, letting the filter automatically delete messages was not accepted by the great majority of the participants. In case of the relatively low-risk situation of using the art recommender, most participants did not let the system choose artworks for them. In both studies, even if participants report trusting a filter, they might not actually want to rely on the system. This issue has to be taken into account in further study and system evaluations; reported trust does not necessarily mean actual usage behaviour will occur that would indicate this trust. Training of the system by the users by correcting it or rating how interesting certain information items are, does appear to increase trust. It does however not necessarily lead to less controlling behaviour. Participants in both the spam filter and recommender study thought training the filter improved it, but this didn't necessarily mean they handed over control to the filter after they observed it making fewer mistakes. The benefits of user-adaptivity, training the filter, and seeing it improve, appear less important to users in their final decision to accept system decisions, than overall potential risks and consequences of using of the filter.

This supports the findings of Cortellessa et al (2005), who argue that focus should be on developing an efficient mixed initiative approach, instead of completely autonomous systems, and users should be offered an appropriate level of transparency of an adaptive system's actions. Additionally, authors such as Kaber and Endsley (2004) describe the dangers associated with handing over control completely: full automation potentially hurts situation awareness, monitoring performance and failure detection and might not result in a decline in perceived workload. The challenge is how to balance benefits of a filter and the risks of its use, how to provide users with appropriate control and training a system to optimal performance, without taking too much time and effort away from the user's primary tasks.

Even though previous literature does state transparency increases trust and acceptance, the art recommender transparency study shows that just showing any transparency feature does not necessarily help trust and acceptance. This particular

transparency feature did not appear to help the user understand the system, illustrating that designing a useful transparency feature is not an easy task (as is also discussed by Herlocker, 2000). In a crisis situation weighing the costs and benefits of a transparency feature will be challenging, but even more important. A relationship between the system and user has to be built before a crisis occurs. As training a filter during a crisis situation is not a very viable option, training should probably occur offline such as in simulated training sessions. This way, in case the filter fails no immediate consequences occur and the user has the chance to get to know the system in a less stressful situation. Raising awareness of filter activity and its adaptivity features in the user is not trivial. Interface choices matter, as illustrated by the participant in the spam filter study who confused a filter train button with a delete button, not realising she was (inaccurately) training a filter. Participants in the spam filter study also could not always accurately report filter settings. In the second study involving the art recommender, interface items such as the additional 'why?' button were not always noticed or used. Lack of awareness might not appear very serious in the case of using a spam filter or an art recommender. In crisis situations however, not knowing whether a filter is active or how to use it, could lead to uninformed choices, increased risk and difficulty in recovering in case of filter mistakes.

Introducing user-adaptive information filters in crisis management practice is challenging. Ways for the user to manage the risk associated with their use have to be devised before they will be accepted. A longer-term relationship has to be built in a setting where the user keeps control over the system but can still benefit from efficiency gains from using it. Perceived benefits of use of a filter have to outweigh the perceived efforts required from users to understand, train and correct them. Transparency has potential to help users make informed choices and develop appropriate trust and acceptance. Producing explanations for users that help them understand the complexities of a filtering system's decisions during a crisis, that also are concise and understandable will not be a trivial matter however. Until it is completely clear how to design a transparent system, careful instruction on such a filtering system, explaining the users on how a system works and how to use and correct could be crucial. The implementation of a user-adaptive information filtering system in emergency management will thus require a lot of user effort before it can actually be used. It appears that trust will remain playing an important role in usage as full system transparency appears difficult to achieve; crisis situations are by nature uncertain and complex, and spending extensive effort and time deciding whether to use a system is impractical. The studies discussed here are steps towards a fuller understanding of these issues, but also show that more research into the relations between acceptance, trust and transparency is necessary.

Acknowledgments. This research is funded by the Interactive Collaborative Information Systems (ICIS) project nr: BSIK03024, by the Dutch Ministry of Economical Affairs under contract to the Human-Computer Studies Laboratory of the University of Amsterdam. The CHIP system is developed by the CHIP (Cultural Heritage Information Personalization - www.chip-project.org) project, part of the CATCH (Continuous Access To Cultural Heritage) program funded by the NWO (Netherlands Organisation for Scientific Research).

References

1. Cortellessa, G., Giuliani, M.V., Scopelliti, M., Cesta, A.: Key Issues in Interactive Problem Solving: An Empirical Investigation on Users Attitude. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 657–670. Springer, Heidelberg (2005)
2. Dzindolet, M.: The role of trust in automation reliance. Int. J. of Human-Computer Studies 58(6), 697–718 (2003)
3. Fogg, B.J., Tseng, H.: The Elements of Computer Credibility. In: Proc. CHI 1999, pp. 80–87. ACM Press, New York (1999)
4. Fogg, B.J.: Prominence-Interpretation Theory: Explaining How People Assess Credibility Online. In: In Proc. CHI 2003, ACM Press, NewYork (2003)
5. Hanani, U., Shapira, B., Shoval, P.: Information Filtering: Overview of Issues. Research and Systems, User Modeling and User-Adapted Interaction 11(3), 203–259 (2001)
6. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proc. CSCW 2000, pp. 241–250. ACM Press, New York (2000)
7. Höök, K.: Steps to Take Before Intelligent Interfaces Become Real. Interacting with computers 12(4), 409–426 (2000)
8. Jameson, A.: Adaptive Interfaces and Agents. In: Jacko, J.A., Sears, A. (eds.) Human-computer interaction handbook, pp. 305–330. Erlbaum, Mahwah, NJ (2003)
9. Jian, J.Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated systems. Int. J. of Cognitive Ergonomics 4(1), 53–71 (2000)
10. Jøsang, A., Lo Presti, S.: Analysing the Relationship between Risk and Trust. In: Proc. International Conference on Trust Management, pp. 135–145. Springer, Heidelberg (2004)
11. 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 ergonomic science 5(2), 113–153 (2004)
12. Meissner, A., Wang, Z., Putz, W., Grimmer, J.: MIKoBOS - A Mobile Information and Communication System for Emergency Response. In: Proc. ISCRAM 2006 Van de Walle, B., Turoff, M. (eds.), Newark, NJ, USA (2006)
13. Muir, B.M., Moray, N.: Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics 39(3), 429–460 (1996)
14. Parasuraman, R., Miller, C.: Trust and etiquette in high-criticality automated systems. In: Proc. CHI 2004, pp. 51–55. ACM Press, NewYork (2004)
15. Pu, P., Chen, L.: Trust Building with Explanation Interfaces. In: Proc. IUI 2006, pp. 93–100. ACM Press, NewYork (2006)
16. Sinha, R., Swearingen, K.: The Role of Transparency in Recommender Systems. In: CHI'02 extended abstracts on Human factors in computing systems, pp. 830–831. ACM Press, New York (2002)
17. Venkatesh, V., Morris, M., Davis, G., Davis, F.: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27(3), 479–501 (2003)
18. Wærn, A.: User Involvement in Automatic Filtering: An Experimental Study. User Modeling and User-Adapted Interaction 14(2-3), 201–237 (2004)