
Interaction with User-Adaptive Information Filters. Trust, Transparency and Acceptance.

Henriette Cramer

Human-Computer Studies Lab
University of Amsterdam
Kruislaan 419
1098 VA, Amsterdam
The Netherlands
hcramer@science.uva.nl

Abstract

This PhD-project investigates interaction with user-adaptive systems. Experiments and user studies are used to explore the factors that lead to trust and acceptance of such systems. This research aims to inform design of transparent user-adaptive and (semi-)autonomous systems. Focus is on interaction with content-based user-adaptive information filters.

Keywords

User-adaptive systems, information filtering, transparency, trust.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. H1.2. Models and Principles: User/Machine Systems.

Introduction

'Intelligent', adaptive technologies are becoming more and more common. Adaptive systems are designed to adapt themselves to the user and often over time reach a certain degree of autonomy. However, users often do not trust adaptive systems to make decisions for them. Major questions in designing interaction with user-adaptive systems and achieving appropriate trust and

Copyright is held by the author/owner(s).

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

ACM 978-1-59593-642-4/07/0004.

reliance are yet unsolved. It is unclear how user interaction with these systems can be optimised. This project takes a closer look at user interaction with user-adaptive systems. Focus is on interaction with content-based, user-adaptive information filters that learn using explicit user feedback. User-adaptive filters adapt informational content to user needs on the basis of user feedback. The dialogue between such a user-adaptive information filter and the user is extremely important in achieving both adequate system performance and acceptance in the user. The system needs to build trust in the user; users have to perceive they can rely on filtering results and hand over control to the system when appropriate. At the same time the system relies on user involvement in providing feedback to the system and improve its performance. The perceived benefits of using the system need to outweigh the burdens of providing this feedback, and also need to outweigh perceived risks of system mistakes. The user needs to keep on using the system so it can learn, even though a system might not directly provide optimal results (as noted by e.g. Höök [3], Hanani [1], Waern [9]). How to manage user perceptions and achieve appropriate user control and trust are yet open questions. This project tries to establish how user-adaptive filters can build appropriate trust and acceptance in their users. Using (experimental) user studies factors are studied that influence trust and acceptance in various contexts. Focus is on risk perception, user control and transparency.

The main contributions of this project will be:

- Theory and interaction models of transparency, trust and acceptance of user-adaptive systems.

- Design guidelines for user-adaptive systems in various usage contexts.

Background: acceptance, trust and transparency

The attitude of users towards a system influences whether a system is accepted and used [8]. Trust in a system also plays a role in acceptance and use [4]. Users need to be willing to depend on the system not to filter out important information and have faith in the notion that adaptive filters get better over time. Jøsang and Lo Presti [5] 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'. Building a trust relationship between a user-adaptive system and users is often difficult, as users cannot always easily inspect or correct the user model the system has built. Machine learning methods for example, are often based on classification models that are not very comprehensible to the user. In previous studies it is suggested that making an adaptive system more transparent to the user could lead to increased trust and acceptance and increases in performance (e.g. [2], [6], [9]). Transparency is one of the factors in trust and acceptance of user-adaptive systems addressed in this project. What type and levels of transparency are needed in which context to achieve optimal use of a system is unclear. For example, Waern [9] shows that the perceived quality of a profile used in automatic information filtering does not necessarily correlate with the profile's filtering performance. Herlocker [2] shows that some transparency types might actually hinder acceptance. The effort needed from the user to build an accurate mental model of an adaptive system needs to be minimal, so the efficiency

gains of using that system are not diminished. This PhD study focuses on making the internal decision making process of information filters based on AI-techniques more transparent and investigating the effects on user behaviour, attitude and acceptance.

Methods

(Experimental) user studies are used to gain more insight in user interaction with adaptive systems. A study into usage of user-adaptive spam filters and an experiment on the effects of offering transparency on acceptance of a content-based recommender system have been completed. These are described first, planned lab experiments are discussed afterwards.

Study: interaction with adaptive spam filters

A study into use and acceptance of adaptive spam filters (as an example of widely available adaptive filters) has been carried out to explore the ways people use and train adaptive filters. Thirty-one participants took part using their regular work email and filter settings. Observation, open-ended interviewing and a questionnaire were combined to study users' behaviour, acceptance and trust in usage and training of learning spam filters. While participants verbally reported quite high levels of trust in their filter, in practice very few of them actually seem to trust a (trained) filter to automatically filter and delete messages. Another finding was that training behaviour seemed influenced by the user's understanding of the (learning) mechanisms of the filter. Some participants for example reported that they did not correct the filter when missed spam messages were very similar to legitimate emails. They feared that if they would mark these specific spam messages the filter would start marking more legitimate mail as spam (false positives). These

similar messages are however most informative to the filter, since they can provide information on the finer distinction between spam and legitimate email. This is one of the examples where an incorrect mental model of a system leads to sub-optimal use.

Experiment: transparency and trust

To gain a deeper understanding of the circumstances in user-adaptive information filters where a user's mental model and transparency can make a difference, an experiment was set up. The experiment investigated the influence of system transparency on user trust in and acceptance of a content-based art recommender system. A system recommending art works on the basis of explicit user rating of other artworks was used (www.chip-project.org). A between-subjects experimental set-up exposed 60 participants to three different versions. Version 1 was not transparent. Version 2 explained why an artwork was recommended. Version 3 stated how sure the system was of a recommendation. Qualitative observation and interview data and quantitative questionnaire data were used in analysis. Findings indicate that offering explanations influences acceptance of single recommendations the user already likes. Transparency in this case was not found to influence users' trust in the recommender system overall. Offering certainty ratings did not have any positive effects. A number of contextual factors influenced the effect of transparency. For example, the affective relationship with art of subject, the possibility of direct evaluation of results (participants could directly assess whether recommendations were accurate and whether they agreed with system explanations) and design of the transparency features, all appeared to affect user attitudes and the effects of transparency.

Planned experiments

The first experiments of this project provided information on interaction, acceptance of and trust in user-adaptive content-based information filters. The need for further information on the effects of transparency on trust and acceptance in various circumstances will be addressed in the remainder of this PhD-project. Future experiments will address effects of transparency on attitudes towards communication filtering in more high-risk situations, where issues of trust and control may be more pertinent. Proposed experiments include an experiment focusing on the effects of transparency in situations where an adaptive information filter is filtering communication between team members. It will focus on trust, acceptance and usage behaviour and study what the effects are on communication between team members and their perceptions of the situation, risks and other team members. A second study will focus on the effects of transparency on risk perception in various contexts (different levels/types of risks), trust, acceptance and effects on for example users' feedback to the adaptive system. The results from these studies are expected to add to theory on transparency, trust and acceptance and guidelines for the development of transparent adaptive information filters and other adaptive systems.

Acknowledgements

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 program funded by the NWO (Netherlands Organisation for Scientific Research).

References

- [1] Hanani, U., Shapira, B., and Shoval, P. Information Filtering: Overview of Issues, Research and Systems. *User Modeling and User-Adapted Interaction 11*, 3 (2001), 203 - 259.
- [2] Herlocker, J.L., Konstan, J.A. and Riedl, J. (2000), Explaining collaborative filtering recommendations. *IProc. 2000 CSCW*, 2000, ACM Press, 241-250.
- [3] Höök, K. Steps to Take Before Intelligent Interfaces Become Real. *Interacting with computers 12*, 4 (2000), 409-426.
- [4] Jian, J.Y., Bisantz, A.M. and Drury, C.G. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics 4*, 1 (2000), 53-71.
- [5] Jøsang, A., and Lo Presti, S. Analysing the Relationship between Risk and Trust. In *Proc. iTrust 2004*, Springer (2004).
- [6] Pu, P., Chen, L. Trust Building with Explanation Interfaces, In *Proc. of the 11th international conference on Intelligent user interfaces*, ACM Press (2006), 93-100.
- [7] Sinha, R., Swearingen, K. The Role of Transparency in Recommender Systems. In *CHI'02 extended abstracts on Human factors in computing systems*, ACM Press (2002), 830-831.
- [8] Venkatesh, V., Morris, M. G., Davis, G.B., Davis, F. D. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly 27*, 3 (2003), 425-478.
- [9] Waern, A. User Involvement in Automatic Filtering: an Experimental Study. *User Modeling and User-Adapted Interaction 14*, 2-3 (2004), 201-237.