



UvA-DARE (Digital Academic Repository)

The effects of transparency on trust in and acceptance of a content-based art recommender

Cramer, H.; Evers, V.; Ramlal, S.; van Someren, M.; Rutledge, L.; Stash, N.; Aroyo, L.; Wielinga, B.

DOI

[10.1007/s11257-008-9051-3](https://doi.org/10.1007/s11257-008-9051-3)

Publication date

2008

Document Version

Final published version

Published in

User Modeling and user-adapted Interaction

[Link to publication](#)

Citation for published version (APA):

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and user-adapted Interaction*, 18(5), 455-496. <https://doi.org/10.1007/s11257-008-9051-3>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

The effects of transparency on trust in and acceptance of a content-based art recommender

Henriette Cramer · Vanessa Evers ·
Satyan Ramlal · Maarten van Someren ·
Lloyd Rutledge · Natalia Stash · Lora Aroyo ·
Bob Wielinga

Received: 10 May 2007 / Accepted in revised form: 26 June 2008 / Published online: 20 August 2008
© The Author(s) 2008

Abstract The increasing availability of (digital) cultural heritage artefacts offers great potential for increased access to art content, but also necessitates tools to help users deal with such abundance of information. User-adaptive art recommender

H. Cramer (✉) · V. Evers · S. Ramlal · M. van Someren
Human Computer Studies Lab, University of Amsterdam, Kruislaan 419,
1098 VA, Amsterdam, The Netherlands
e-mail: hcramer@science.uva.nl

V. Evers
e-mail: evers@science.uva.nl

S. Ramlal
e-mail: satyanramlal@gmail.com

M. van Someren
e-mail: maarten@science.uva.nl

L. Rutledge
Telematica Institute, P.O. Box 589, 7500 AN, Enschede, The Netherlands
e-mail: Lloyd.Rutledge@cwi.nl

L. Rutledge
CWI, Kruislaan 413, 1098 SJ, Amsterdam, The Netherlands

N. Stash · L. Aroyo
Eindhoven University of Technology, P.O. Box 513, 5600 MD, Eindhoven, The Netherlands
e-mail: natalia.stash@gmail.com

L. Aroyo
e-mail: l.m.aroyo@cs.vu.nl

N. Stash · L. Aroyo
VU University Amsterdam, De Boelelaan 1081, 1081 HV, Amsterdam, The Netherlands

B. Wielinga
Human Computer Studies Lab, University of Amsterdam, Kruislaan 419, 1089 VA, Amsterdam,
The Netherlands
e-mail: wielinga@science.uva.nl

systems aim to present their users with art content tailored to their interests. These systems try to adapt to the user based on feedback from the user on which artworks he or she finds interesting. Users need to be able to depend on the system to competently adapt to their feedback and find the artworks that are most interesting to them. This paper investigates the influence of transparency on user trust in and acceptance of content-based recommender systems. A between-subject experiment ($N = 60$) evaluated interaction with three versions of a content-based art recommender in the cultural heritage domain. This recommender system provides users with artworks that are of interest to them, based on their ratings of other artworks. Version 1 was not transparent, version 2 explained to the user why a recommendation had been made and version 3 showed a rating of how certain the system was that a recommendation would be of interest to the user. Results show that explaining to the user why a recommendation was made increased acceptance of the recommendations. Trust in the system itself was not improved by transparency. Showing how certain the system was of a recommendation did not influence trust and acceptance. A number of guidelines for design of recommender systems in the cultural heritage domain have been derived from the study's results.

Keywords User-adaptivity · Human-computer interaction · Recommender systems · Transparency · Trust · Acceptance · Cultural heritage

1 Introduction

Museums and archives are digitising their material to both preserve and extend user access to cultural heritage content. Combining digitised collections with mobile technology and other innovations offer great potential for improving museum visits and the exploration of cultural heritage in general. This can result in the creation of (virtual) exhibitions where physically dispersed objects are virtually co-located and interconnected. However, such abundance of information can cause users to have difficulty finding relevant and interesting content. A greater variety of users also means that exhibitions might not appeal to all of them. Tailoring content to users offers the prospective of a more interesting and satisfying experience when interacting with cultural heritage content (Bowen and Filippini-Fantoni 2004). User-adaptive recommender systems are one way of offering such personalisation. Recommender systems can help users cope with an abundance of information items by offering those items the user is likely to find interesting (Burke 2002). Recommender systems learn about the user's preferences and build a user profile for each user. These profiles can be based on either similarities between users (collaborative- or social-based filtering), content features of recommended items (content-based), or a hybrid of these approaches (Burke 2002). Based on the user's profile, the system then presents the user with information items that are likely to be of interest. Such recommender systems are rapidly becoming a mainstream feature, offered in a variety of domains. Examples include Amazon.com's book recommender and Pandora.com's personalised radio stations.

The extent to which the user is satisfied with a user-adaptive recommender depends on quality of its recommendations, but also greatly depends on user-system interaction.

Designing the interaction between a user and a recommender system is challenging. A recommender system has to successfully learn from the user, adapt the user's profile and present him or her with interesting items. The user profile has to be adequate and the system's criteria for recommendations have to match the criteria that are relevant to the user and the specific task he or she is trying to accomplish. When developing a recommender system in a cultural heritage context that recommends artworks of interest to the user, the artist might be relevant to one user, while the artwork's theme is not; and vice versa for another user. Especially in the cultural heritage domain it is difficult to offer recommendations as the user may select objects of interest on emotional and aesthetic grounds while systems more easily use descriptive characteristics given by art experts related to for example artists, style, techniques, repositories, dimensions, themes, depicted events and to some extent content and aesthetic features such as colour and shapes. Additionally, criteria are application dependent. Recommending artworks for an entertaining, personalised tour in a museum would require other criteria than providing artworks for learning about art in an educational setting (even though these applications need not be mutually exclusive). As recommenders learn from interaction with the user, they will only achieve optimal performance over time. Therefore, the dialogue between recommender and user needs to build a trust relationship, where the user feels he or she can depend on the system. This is particularly challenging, as recommender systems cannot always be 100% accurate in predicting the user's preferences. It is imperative that the user provides feedback to facilitate the system's learning process and improve recommendations. When explicit feedback is used (such as explicit ratings of items as interesting or uninteresting) the user needs to be persuaded to provide the system with enough explicit feedback to improve the user profile and allow the system to reach a high level of performance. Users' understanding of a system might play a key role in this interaction.

It might be difficult for users to assess whether a system's goals and reasoning processes are in line with their own interests and reasoning about e.g. art objects. A transparent system instead allows the user to understand the way it works and explains system choices and behaviour. Understanding a personalisation system might additionally improve interaction (see e.g. [Tintarev and Masthoff 2007](#); [Waern 2004](#)). When a user has an inaccurate mental model of how the recommender works, he or she may provide inappropriate feedback to the system ([Waern 2004](#)). Better understanding of a system can help users decide whether they can trust a system ([Lee and See 2004](#)) and can influence user attitudes towards a system ([Alpert et al. 2003](#)). Increasing transparency of user-adaptive systems could thus increase trust and acceptance of such systems ([Herlocker et al. 2000](#); [Sinha and Swearingen 2002](#); [Höök 2000](#)).

The study presented in this paper further investigates whether transparency increases users' trust and acceptance of recommender systems. The structure of this paper is as follows: Section 2 reviews related literature on interaction with user-adaptive systems, trust, acceptance and the importance of transparency and outlines the hypotheses of the study. Section 3 describes the methodology of the reported study. This section also describes the CHIP cultural heritage recommender system used in this study. Section 4 reports the study's quantitative results and its qualitative observations. Section 5 further discusses these results and provides an overview of lessons

learned about users' interaction with art recommenders. This section also discusses the limitations of this study. Section 6 ends this paper with the study's conclusions.

2 Related literature and hypotheses

2.1 Personalisation in the cultural heritage domain

The advancement of information and communication technologies plays a significant role in exploring new ways of interaction between museum collections and their (online) visitors. Specifically the evolution of the Web as a major communication medium offers increased access to large and dispersed sets of artworks. This increased access to cultural information also means that cultural heritage visitors might need help to deal with information overload. Additionally, such wider access to information about artworks opens the possibility for a larger and more varied audience to access cultural heritage information. This audience will likely consist of multiple types of users, who do not all have the same needs (Wubs and Huysmans 2006). This presents the challenging task of finding out what the goals, desires and preferences of such users are, and how to adapt correspondingly and present personalised museum collections to visitors and (online) users.

Digitised collections offer the possibility to adapt to the individual user. A decade ago, Picard (1997) identified the need for such personalisation of online museum collections. User-adaptivity and personalisation can help present the individual user with information tailored to his or her specific needs. Personalisation can help overcome problems of information overload when the size of a collection is overwhelming to the visitor. It can also help cater for a more diverse public when one exhibition would not suffice for different types of users. Personalisation additionally has the potential to improve user interaction with digitised collections by supporting user navigation and providing assistance in finding appropriate and interesting information. It can also serve educational goals when information about personal characteristics, such as age, education and familiarity with a collection, are used to support better comprehension of a collection. Bowen and Filippini-Fantoni (2004) also argue for a more personalised experience. They claim personalisation enables change in the museum's mass communication monologue paradigm and turn it into an interactive dialogue in which information is exchanged. Personalisation thus could help both learn from and adapt to online and physical visitors.

Various studies have explored the possibilities of both increasing access to cultural heritage content and personalising content to individual users. Such endeavours can for instance take advantage of the domain's extensive availability of descriptions of characteristics of cultural items and other metadata. Experts at cultural heritage institutions often provide extensive data on the artworks they manage. This enables the use of content-based techniques for personalisation and for presenting new strategies to search and navigate collections. Various information retrieval and search techniques can support access to cultural heritage content. Semantic Web techniques have been used in multiple projects. The MuseumFinland Project (Hyvönen et al. 2005) is an example of using Semantic Web technologies to provide multi-dimensional access to

cultural heritage collections in Finland. The main goal of the project is to provide a global view to distributed collections as one ‘seamless’ national (virtual) collection, the “Museum of Finland”. It offers intelligent services to users for searching and browsing the collections, as well as easy content publication for museums. It also offers possibilities for personalised access to the collections. Other authors who describe usage of Semantic Web technologies for adapted presentation of cultural heritage content include [Schreiber et al. \(2006\)](#) and [Sinclair et al. \(2006\)](#). These authors focus on information retrieval and taking advantage of the knowledge-richness of the cultural domain. They use common ontologies to allow easier access to diverse collections and for preservation of the relationships between information items. Another example of a project using Semantic Web techniques is the CHIP project ([Aroyo et al. 2007](#)). The CHIP recommender system is used in this study.

The new diversity of museum audiences means that not only experts’ concerns have to be reflected. The annotations and other content information about artworks used to personalise access to cultural heritage do not necessarily have to be supplied by museums and art experts alone. [Trant \(2006\)](#) for example describes how users can tag items and provide their own descriptions and keywords. These tags can be used to understand patterns in museum visitors and to explore how expert and non-professional vocabularies differ. [Carmagnola et al. \(2008\)](#) describe how these tags can also be exploited to improve recommender systems. To provide a personalised experience, it is crucial to elicit accurate user characteristics and align them with the terminology and structures the domain experts in museums use to describe their collections. When such concerns are met, personalisation can offer many new opportunities in presenting cultural heritage information.

Adaptive techniques have been used to generate personalised presentations about artists and art collections. Personalised presentations combine information items and artworks to provide a narrative that is tailored to the individual user. The ArtEquAKT project ([Kim et al. 2006](#)) is a project that focuses on using natural language technologies to automatically extract relevant information about the life and work of artists from online documents. This information is then presented to the user in personalised artist descriptions in the form of biographical narratives. These are adapted to the user’s knowledge level. This project illustrates that it is possible to generate user-adapted narratives by extracting and structuring information from knowledge bases using ontologies designed for this domain. The Multimedia Casus project ([Wang 2005](#)) is another example of tailor-made presentation generation and using concept structures for navigating art collections. The project presents a collection of artworks of a performance artist together with artwork-related information used for preservation and re-installations. The Multimedia Casus project proposes an infrastructure for personalised presentations and personalised guided tours for media-rich collections of culture heritage items. These projects show that ways to generate personalised narratives are feasible and available.

Various studies have explored the application of personalisation techniques in supporting physical visits to museums. [Proctor and Tellis \(2003\)](#) concluded from a 2002 trial with a handheld tour in London’s Tate Modern that visitors enjoyed tours more when they were adapted to their interests and abilities. Another example of a mobile guide system is provided by [Oppermann and Specht \(1999\)](#). They also found

encouraging results on the usefulness of such a personalised guide, be it through expert evaluation and commentary. The PEACH project (Rocchi et al. 2004; Goren-Bar et al. 2006; Stock et al. 2007) supplies users with personalised presentations of cultural content on mobile devices in museums. It offers animated agents that help motivate visitors and focus their attention and post-visit summaries that reflect individual user interests. An evaluation with 110 museum visitors offered encouraging results, concluding that major components of their personalisation technology were suitable for use by the general museum public (Stock et al. 2007). Another promising evaluation of a virtual character guiding exhibition visitors is reported by Damiano et al. (2008). Beyond guiding the user through an exhibition, adaptation and context-awareness can also be applied for changing the user's exhibition environment itself. The LISTEN project for example generates immersive augmented audio presentations tailored to the user's context (Zimmermann and Lorenz 2008). These projects show the potential of personalisation techniques in real applications in cultural heritage settings.

The need for user-adaptivity and encouraging results from earlier projects show that personalisation and user-adaptive applications are relevant in a cultural heritage context. Personalisation offers opportunities in overcoming information overload, in generating personalised presentations and adapting to the needs of the specific visitor. However, personalisation also brings challenges for interaction, which have to be addressed. Such issues are discussed in the next section.

2.2 Interaction with user-adaptive systems

User-adaptive systems have been suggested as a way to improve human-computer interaction in situations where different users might have different goals, needs and knowledge. User-adaptive systems can adapt to the specific user's needs and change with changing user interests (Benyon 1993; Brusilovsky 1996). However, apart from the advantages such adaptivity might bring, user-adaptivity also provides new challenges for interaction design. Jameson (2003) provides an overview of issues related to user-adaptivity, such as controllability, privacy, obtrusiveness, breadth of the user's experience, predictability and transparency. User-adaptive systems take decisions on behalf of the user; this influences controllability of a system and might conflict with the user's need to control a system's actions directly. User-adaptive systems use data about the user. A concern might be that this information is used in undesirable ways, affecting the user's privacy. Obtrusiveness might be an issue when the adaptive system itself requires attention from the user, taking away attention from the user's primary task. User-adaptive systems that take over tasks from the user can also affect the breadth of the user's experience. For example, a user might not learn as much about a domain when information is filtered for him or her and the user may inappropriately rely on the information provided by the system alone. Höök (1997) also notes that user-adaptivity might conflict with usability principles such as making a system transparent, understandable and predictable. Adaptive systems might change and adapt in unpredictable ways without the user being able to understand why the system behaves in a certain way. User understanding of the system can affect both the user's attitude towards the system and the user's feedback. When the system behaves in ways the

user does not expect (e.g., recommending something he or she does not like or need), the user may find the system unpredictable and unreliable (Höök 1997). If the user has an inaccurate mental model of the way the system learns, he or she might also change his or her feedback towards the system in such a way that performance decreases (e.g. Waern 2004). All issues above have to be addressed in designing users' interaction with a user-adaptive system.

Different contexts and types of users might ask for different types of adaptivity and user control. Alpert et al. (2003) note that a personalisation technique that might be considered desirable by users in one context might not be met with the same enthusiasm in other contexts. Alpert and colleagues studied user interaction with a user adaptive e-commerce web site for computer equipment sales and support. In Alpert's study participants had mixed reactions to the collaborative filtering feature. Users felt positive about the system that presented new products and services related to the users' current search activities and their recent past activity on the website. However, "some participants were sceptical of systems that try to predict a user's current goals or needs and adapt content accordingly based on implicit information from some disjoint time in the past". Alpert and colleagues found that participants reacted positively to the theoretical possibility of such a feature, but that they did not expect such features to actually do well in practice. They noted that in other e-commerce settings such as movie or book recommendations, users appear to have no objections towards collaborative techniques. Not only the application context, but also characteristics of users themselves might influence acceptance of adaptive systems. Specifically in the domain of user interaction with personalised museum guides, Goren-Bar et al. (2006) found that personality traits relating to the notion of control have a selective effect on the acceptance of the adaptivity dimensions. One of their findings was that users who feel less in control of their own beliefs and actions prefer to maintain control over a system and prefer non-adaptivity. Jameson and Schwarzkopf (2002) investigated the level of controllability of a user-adaptive system, and performed a study in which users got to determine the nature and timing of adaptations. Jameson and Schwarzkopf concluded that the preferable form of control and automation for a task depend on factors ranging from individual differences and system properties that are relatively stable to unpredictable situational characteristics; there is no one level that fits all situations. Zimmermann and Lorenz (2008) for example claim that focusing on control over an adaptive art-related system might be contradictory to building a relationship with the presented art.

Which adaptations and level of control are suitable have to be determined for a particular context, application and the specific user. This needs to be reflected in the user-system dialogue. This dialogue between the user-adaptive system and user appears crucial in achieving both excellent performance and user satisfaction. In the context of user-adaptive recommender systems, this dialogue should provide the user with high-quality recommendations and information. At the same time, the dialogue needs to provide the system with information to improve its recommendations. The user should be persuaded to invest time and effort to provide the system with enough useful training data to improve its recommendations. As Benyon (1993) points out, user preferences for information items can change over time. This means that a user's profile that might have been suitable at one time, might not always be suitable at other

occasions. This might necessitate continuous adaptation to the user's feedback, while taking into account the issues described above, such as controllability, obtrusiveness and user understanding. User-adaptive systems raise the expectation to 'know what the user wants' and to make (semi)-autonomous decisions on what is relevant to the user. The system's user profiling has to be adequate to process the user's feedback and the criteria used for recommending have to be suitable for the user's task. The user on the other hand has to be willing to release control, accept the system and delegate tasks. In the cultural heritage domain, users have to feel confident they can let a system for example adapt a museum tour for them, or search out those artworks that would be interesting to them. The dialogue between system and user should overcome challenges inherent to user-adaptivity and instil trust in the user that they can indeed let a system make decisions for them in their specific situation.

2.3 Trust in user-adaptive systems

Reeves and Nass (1996) have shown that people tend to interact with computer systems as if they were social actors. Users' affective and social reactions to automated tools affect acceptance of and interaction with such systems. A possible implication is that social rules and principles of human-to-human interaction can also be applied to human-computer interaction. This includes the concept of trust. Trust is thought to be one of the important affective elements that play a role in human interaction with technology (Lee and See 2004; Dzindolet et al. 2003). Making a system reliable is not enough to achieve appropriate delegation to that system (as observed by e.g. Parasuraman and Miller 2004). Trust also depends on the interaction with the user. Users need to adapt their levels of trust in order to optimise human-machine performance and benefit fully from the cooperation (Rajaonah et al. 2006a, b, 2008). Users may have a level of trust in a device, but also have a level of trust in the success of cooperation with the device (see also Muir 1994; Muir and Moray 1996). Even if an appropriate level of automation can be determined, there is no guarantee that users will trust a system to an appropriate degree and will actually decide to use it. In the context of adaptive cruise control assisting drivers, Rajaonah and colleagues found that trust in the cooperation with a device determines the quality of interaction between the user and system, as well as appropriate use. In an experiment using a flight simulator and an engine monitoring system that provided advice on engine fault diagnosis, Parasuraman and Miller (2004) found that user performance and trust were lowered by poor automation etiquette, regardless of reliability. In their study a non-interruptive and patient communication style, instead of using interruptive and impatient worded system messages, increased trust and performance. Even though a high-reliability, good etiquette system version yielded the best user performance and the highest trust; they found that good automation etiquette can compensate to a certain extent for low automation reliability. Performance alone does thus not completely determine whether users take full advantage of potential system benefits.

Lee and See (2004) describe how trust guides reliance on automation, even while it does not completely determine it. They provide a comprehensive review of trust from various perspectives, including organisational, sociological and interpersonal standpoints. According to Lee and See, trust affects reliance as an attitude rather than as a

belief, intention, or behaviour. Certain beliefs, for example about the system, determine trust and trust in turn might result in certain intentions and behaviours. Lee and See state that trust especially guides reliance when a complete understanding of a system is impractical. They do however note that reliance on automation is also determined by the operating context and goes beyond trust alone. Dzindolet et al. (2003) note that users do compare their own abilities with those of a system. On the basis of the outcome of this comparison, users decide whether or not to rely on a system. Authors such as Castelfranchi and Falcone (2000) and Jøsang and Lo Presti (2004) also note that even though a user might trust a system, he or she might not choose to rely on it when the user perceives the risks as too great.

In this study we apply the definition of trust by Jøsang and Lo Presti (2004): “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”. A similar definition is offered by Lee and See (2004), who define trust as “the attitude that an agent will help achieve an individual’s goals in a situation characterized by uncertainty and vulnerability”. Parasuraman and Miller (2004) state that trust is users’ willingness to believe information from a system or make use of its capabilities. This definition is related to an alternative concept proposed by Fogg and Tseng (1999), who introduce the concept of credibility. Credibility (synonymous to believability) is proposed to consist of trustworthiness (the perception that the source has good intentions) and expertise (or competence) of a system. In this paper, the concept of trust consists of trust in the intentions of a system (goal alignment) and trust in the competence of the system. Competence is seen as the perceived skill of the system: the extent to which it is able to offer the right recommendations. The perception of the alignment of goals of the system and the user’s goals, coupled with a belief that a system will perform its task competently, form the basis of trust. In the context of this paper, trust refers to the user’s willingness to depend on a system and its recommendations in the specific context of the user and his or her task(s), even though the system might make mistakes.

Trust is of specific interest in interaction with user-adaptive systems. A number of factors might play a role in the perceptions users have of the goals of the system and the intentions of its developers and other users. Briggs et al. (2004) note that personalisation itself plays a role in trust formation. In the context of e-commerce websites, in an experiment using a (mock) website offering travel insurance, Briggs found that participants felt more positive about the site when it offered a personalised quote based on participants’ circumstances. They felt, for example, that they had been offered more choice, that the site was more predictable and that it was easier to use. Such perceived benefits of personalisation might however be dependent on whether the intentions and capabilities of a personalisation system instil trust in the user. In some situations, trust may be a prerequisite to get the personal information needed to personalise a system, invoking privacy concerns (Langheinrich 2001). In recommender systems, trust in the information sources a system uses to generate its recommendations can also play a role in whether a user trusts the system. Authors such as Wang and Benbasat (2005) and Xiao and Benbasat (2003) studied trust in online e-commerce recommender settings. In such settings trust in the online vendor’s goals in providing recommendations might also play a role. In collaborative-based

recommenders, users need to assess whether they trust the intentions and actions of other users (Victor et al. 2008). In addition, in contexts where a system might provide information about a recommended item not immediately verifiable by users, trust in the correctness and credibility of the information might play a role as well (Fogg and Tseng 2003). For example, when a recommender system provides recommendations based on how comfortable a certain product will feel, users cannot know whether this is correct before they actually buy and feel the product.

Perceptions of competence of a personalised system can be influenced by issues inherent to adaptivity. User-adaptive systems might not offer a high level of performance from the start. If performance is low in the beginning, users' trust levels should start out low as well, and gradually increase as training progresses and performance increases. However, trust rapidly decreases when users notice errors and only slowly increases as a system performs without errors (Dzindolet et al. 2003; quoting Muir 1994; Muir and Moray 1996). This makes achieving trust difficult in adaptive systems that have not been pre-trained. Lee and See use the terms calibration, resolution and specificity of trust to describe such (mis)matches between trust and the capabilities of automation. Calibration refers to the correspondence between a person's trust in a system and the system's capabilities (see also Lee and Moray 1994; Muir 1994). Resolution of trust refers to what extent such changes in system capability affect trust. Trust also has specificity; trust can be more or less specific to part of a system or to changes in a system's capability. Getting these aspects of trust right is especially challenging in a personalised system, as the nature of these systems implies that due to adaptation to e.g. user feedback the system's capabilities change; the specificity and resolution of trust in the user might not match these changes.

In the context of a recommender in the cultural-heritage domain, users need to be sure that the purpose of the system matches their own and that the system competently recommends artworks that are of interest to them. Furthermore, when they are not familiar with the sources that provide the information, they might need to trust that the information provided about the artworks is trustworthy. Moreover, if multiple sources are combined to generate integrated personalised content for a user (e.g. Sinclair et al. 2006), it might be even more difficult for the user to trust the information and the system. However, it has to be taken into account that trust toward recommenders in a cultural heritage context is different than trust in for example an e-commerce context. In most e-culture contexts, inaccurate recommendations do not pose a great risk to the user. When users trust a system to generate a personalised tour from recommended artworks and it turns out the recommended tour was not suited to their needs, they only risk a less entertaining or less educational visit. They are not likely to suffer serious consequences, whereas in commercial settings an inaccurate recommendation may result in financial loss. However, acceptance of the system is still likely to be guided by user trust and perceptions of the goals and competence of the system.

2.4 Combining trust in and acceptance of user-adaptive systems

The section above describes how trust guides reliance or dependence on a system. Reliance is related to acceptance and use of a system. The Technology Acceptance

Model (Davis 1989; Venkatesh et al. 2003) describes a number of concepts that influence acceptance of a system. According to the model (updated by Venkatesh et al. 2003), performance expectancy, effort expectancy (how much effort it will take to learn and use a system), social influence and facilitating conditions (whether the user has enough time and resources to use the technology) are expected to influence intent to use a system and actual usage behaviour. Several authors have studied possible extensions to the Technology Acceptance Model (see e.g. Ndubisi et al. 2005; Ma and Liu 2004; Kloppe and McKinney 2004; van der Heijden 2004). van der Heijden (2004), for example, showed that the purpose of a system plays an important role in determining the predictive importance of respective elements of the technology acceptance model. Perceived usefulness is important in achieving user acceptance when interacting with work- and task-oriented systems. If users interact with 'hedonic' systems that are aimed at fun and entertaining the user instead, perceived enjoyment and perceived ease of use are more important than perceived usefulness. Van der Heijden suggests that focusing on the type of system could increase understanding of technology acceptance, in addition to extending acceptance models with more factors that influence system acceptance.

Trust has been shown to affect system acceptance and has been combined with the technology acceptance model, by such authors as Wu and Chen (2005); Wang and Benbasat (2007); Pavlou (2003) and Gefen et al. (2003). These authors did so mostly in the context of e-commerce and on trust in services. They primarily focus on trust in online transactions and the relationship between a buyer and an online vendor. Integration of acceptance models with other trust-related aspects, such as trust in a system itself, in other contexts is still scarce. In a non-commercial e-culture context, such a consumer-buyer relationship is arguably absent. The trust building process might therefore be more focused on whether the system goals match the users' goals and whether it would be competent in giving recommendations. Trust in the intentions of the recommender might have less of an influence. This study thus focuses on the trust a user has in the system itself and the delegation of tasks to that system. System and recommendation acceptance in this study refers to whether the system and its recommendations actually would be used.

2.5 Transparency of user-adaptive systems and study hypotheses

Maes (in a discussion between Shneiderman and Maes (2007)) states that one of the great interaction design challenges is how to achieve understanding and control in user-agent interaction. If the user understands how a system works and can predict system actions and outcomes, the user can focus on his or her task instead of trying to figure out the system. In the case of user-adaptive recommender systems, it may be even more imperative for the users to understand how the system decides on a recommendation. In order for the system to reach an optimal level of performance, it needs to learn from the users' implicit or explicit input, for example by analysing user's search behaviour or by receiving explicit feedback on recommendations. When a user is not aware of or does not have an accurate understanding of how the system makes decisions, the user will not 'train' the system properly and it may not be possible

for the system to improve. Better understanding of the system could help improve such aspects of interaction. Transparency aims to increase understanding and entails offering the user insight in how a system works, for example by offering explanations for system behaviour. This study investigates the effects of transparency on trust and acceptance; its hypotheses based on prior research into transparency are listed in this section.

System transparency may not always improve user interaction. It is not always possible for the user to construct an accurate mental model in a reasonable timeframe. Fogg (2003) notes that in the context of credibility, the impact of any interface element on users' attitudes depends on to what extent it is noticed (prominence) and what value users assign to the element (interpretation). Parasuraman and Miller (2004) notes that as automation gets more complex, users will be less willing and able to learn about the mechanisms that produce the automation's behaviours. Höök (1997) notes that it is not necessarily desirable to have a system explain how it works in full detail because the full details might be alienating to a layman user. It might not be necessary to explain all details to achieve adequate understanding of the system's adaptivity. Waern (2004) shows that users cannot always recognise good quality user profiles and cannot always improve them when given the opportunity. Besides the effort needed from the user to process transparency information, other adverse effects have to be considered as well. Herlocker et al. (2000) found that poorly designed explanations can actually hinder the acceptance of individual recommendations. Dzindolet et al. (2003) found that explaining why errors might occur increased trust in automated tools, even when this was not appropriate. Therefore, transparency and task performance can influence each other negatively (see also Cheverst et al. 2005). Bilgic and Mooney (2005) note that explanations should not aim just to promote ('sell') a system's recommendations, but that explanations should enable the user to accurately assess recommendation quality. Appropriate implementations of transparency need to overcome these challenges.

In spite of these potential problems, previous authors have emphasised the importance of transparency (Jameson 2003; Höök et al. 1996). Sinha and Swearingen (2002) for instance, evaluated transparency and recommendations for five music recommender systems. Mean liking and confidence were rated higher for recommendations that were more understandable to participants. Other prior studies also suggest that making adaptive systems more transparent to the user could lead to a more positive user attitude towards using a system and increases in system performance (e.g. Kay 2006; Höök et al. 1996; Höök 2000; Carmagnola et al. 2006; Gregor and Benbasat 1999). We expect that transparency will have a positive effect on acceptance of user-adaptive recommender systems.

H1: Users are more likely to accept a user-adaptive recommender system with a more transparent decision making process.

Lee and See (2004) state that appropriate trust and reliance depend on how well the capabilities of a system are conveyed to the user. McAllister (1995), quoting Simmel (1964), notes that "the amount of knowledge necessary for trust is somewhere between total knowledge and total ignorance. Given total knowledge, there is no need to trust, and given total ignorance, there is no basis upon which to rationally trust". Lee and See

argue that promoting appropriate trust may depend on presenting information about a system in a manner compatible with analytic, analogical, and affective processes that influence trust. They identify three types of goal-oriented information that appear to contribute to the development of trust: information on current and past performance (system expertise), information on the system's process (to assess appropriateness of the automation's algorithms for the user's situation and goals) and purpose (to match the designer's intent for the system to the user's goals). Increasing transparency of a system can help users decide whether they can trust the system.

H2: Users will have more trust in a system with a more transparent decision making process.

A number of authors describe systems with various features that aim to support trust and acceptance by making a system's reasoning process understandable and providing insight in system competence. In the context of the Semantic Web, [McGuinness and Pinheiro da Silva \(2004\)](#), describe an explanation system that explains answers from Semantic Web agents that use multiple sources to devise an answer to user questions. Their explanations include information on the origin of answers or how they were derived. Other examples of transparency features can be found in the context of e-commerce product recommendations. [McSherry \(2005\)](#) discusses an approach to finding a fitting product recommendation by guiding the user through the features of the available products, which can for example help understanding the trade-offs and interdependencies between product features. [Shimazu \(2002\)](#) discusses a system mimicking the interaction between a customer and a store clerk to come to a final understandable product recommendation. [Pu and Chen \(2007\)](#) have focused on trust in the recommender based on the user's perception of its competence, and its ability to explain the recommended results. Their organisation interface, which organised recommender results by trade-off properties, was perceived as more capable and efficient in assisting user decisions. [Cortellessa et al. \(2005\)](#) also evaluated the importance of explanations in interactive problem solving systems and found that explanations are needed more when the system's problem solving fails.

In an e-culture context especially, recommendations can be based on a lot of different attributes of artworks (e.g. aesthetics, history, themes, materials, popularity, other people's taste) and could serve a variety of goals (e.g. finding one artwork the user would find aesthetically pleasing, building a tour through a museum, suggesting new artworks the user could learn about). Perception of the recommender's competence is an issue. When the user cannot assess for what goal a system has been built and whether the properties used by the system match his or her criteria, the interaction will not likely be satisfactory. The study described in this paper stems from the premise that transparency of a user-adaptive system that performs well will increase the users' perception of system competence.

H3: Users will perceive a user-adaptive recommender system with a more transparent decision making process as more competent.

Various studies have investigated the effects of different types of transparency features. [McNee et al. \(2003\)](#) found that adding a confidence metric to a movie recommender, indicating which recommendations were 'more risky', increased user satisfaction and

influenced user behaviour. However, they also found that more experienced users of the movie recommender were less satisfied with the system after being instructed about the confidence rating. Wang and Benbasat (2007) compared the effects of three types of explanations on user trust in an e-commerce recommender. The recommender system used in their study advised users what digital camera to buy based on their answers to questions about their product needs (not on the basis of the user's ratings of other products). Wang and Benbasat compared the effects of, in their terminology, 'why', 'how' and 'trade-off' explanations. 'Why' explanations were designed to demonstrate the recommended is designed to fulfil the user's needs and interests. These 'why' explanations justified the importance and purpose of questions posed by the recommender and justified final recommendations. 'How' explanations revealed the line of reasoning used based on consumer needs and product attributes preferences in reaching recommendations. 'Trade-off' explanations offered decisional guidance, helping users to make proper trade-offs among product features and costs of these features (e.g. a digital camera offering more zoom capabilities will also cost more). Wang and Benbasat found that 'why' explanations increased benevolence beliefs; the perception that a recommender acts in the consumer's interests. They found that 'how' explanations increased participants' belief in competence and benevolence of the recommender. 'Trade-off' explanations increased integrity beliefs. Herlocker et al. (2000) compared the effects of 21 different explanations for a movie recommendation on participants' acceptance of the recommendation. The participants had to indicate how likely it was that they would go and see the movie that was recommended by the system. Participants reported that the explanations were important. Histograms of other users' ratings of an item were found to be the most convincing way to explain why a recommendation had been made. Indications of a system's past performance, likeness to similar previously rated items, and domain-specific content features (such as favourite actress) were most compelling in justifying a recommendation. Some other types of explanations, such as correlations between information items, were difficult to understand and actually decreased acceptance of recommendations. Tintarev and Masthoff (2007) provide a number of guidelines for recommender system explanations. They, for example, advise that explanations needs to be tailored to the user and context, as different users might find other criteria important in selecting recommendations.

H4: Understandable explanations of the reasons for a particular recommendation will increase acceptance and trust more than other types of transparency features.

The studies evaluated in this section studied different types of transparency and showed different effects of transparency on user understanding and satisfaction. There is no conclusive evidence that a transparent system is always trusted more and it is yet unclear how different types of transparency can lead to increased trust and acceptance of user-adaptive systems. To take research into transparency and trust in recommender systems further it appears that empirical evidence is needed on the effects of different transparency features on actual user attitudes and behaviour. The study described in this paper explores whether offering transparency of a recommender system's decision making process allows users to understand why certain recommendations are

made and increases their acceptance of a system in an e-culture context. An experimental set-up is used to study the effects of system transparency on user trust in and acceptance of content-based recommender systems that rely on explicit user ratings of art objects. Making the content-based criteria for a recommendation transparent is expected to increase understanding of the system. If users think the criteria the system uses to select recommendations are suitable, users' perceptions of competence, trust and acceptance should also increase.

3 Method

As outlined above, this study aims to further investigate, both quantitatively and qualitatively, the effects of transparency on trust and acceptance of user-adaptive systems. The experiment discussed in this paper had a between-subject design. Its independent variable was transparency (with a transparent condition, a non-transparent condition and a condition that showed information on how certain the system was of its recommendation, but did not offer more transparency of reasons behind recommendations). Dependent variables included perceived competence, understanding, trust in, and acceptance of the content-based recommender system. A user-adaptive, content-based art recommender, the CHIP system (Aroyo et al. 2007; www.chip-project.org) was used for the experiment. An additional goal of the study was to identify possible usability issues of this system. Participants interacted with the system, were observed, interviewed and filled out a questionnaire. In this section the CHIP system will be introduced first, after which the three experimental conditions, measures and procedures used in the study are discussed in detail.

3.1 CHIP system

The CHIP (Cultural Heritage Information Personalisation) system recommends artworks from the collection of the Rijksmuseum in Amsterdam, the Netherlands, based on the individual user's ratings of artworks. The CHIP system uses the semantic annotations of artworks to improve recommendation-based access to these objects. The artworks have been annotated by the Rijksmuseum's staff of experts. These properties include artist (such as Rembrandt), place and time (such as Amsterdam, 1700–1750), themes (such as animals, trompe l'oeil, or musical instruments), genre, materials, techniques, and a number of other content features. The CHIP system offers users the possibility to view and rate Rijksmuseum artworks, such as paintings and sculptures. On the basis of users' ratings of these artworks the CHIP system builds individual user profiles. The CHIP system calculates a prediction of interest for other artworks based on user preference patterns in properties of the artworks that were rated. The recommendation processing is content-based, using positive ratings of artworks and properties to find similar artworks to recommend.

The overall goal of the CHIP system is to personalise access to digitised cultural heritage, in close collaboration with museums. Another goal is to engage and educate the user by offering personalisation. The target audience of the current CHIP projects is anyone who would come to the Rijksmuseum's website or the museum itself—a

broad cross-section of people, but all sharing an interest in art. The CHIP project involves three main applications at this time: the artwork recommender used in this study, a ‘Tour Wizard’ to generate navigational routes through the digital Rijksmuseum collection, and a ‘Mobile Tour’ on a PDA that uses Tour Wizard results to generate personalised tours in the physical Rijksmuseum. Previous research, such as Wubs and Huysmans’s (2006) studies into user goals, interests and navigation patterns of users of cultural heritage collections, as well as van Setten’s (2005) approach for supporting people in finding information, based on hybrid recommender systems and goal-based structuring of information, inspired development of the CHIP system.

3.2 Experimental conditions

The study stems from the premise that system transparency influences understanding, trust and acceptance. From the literature review in paragraph 2.5 it was apparent that different types of transparency are thought to influence understanding, trust and acceptance in different ways (e.g. Herlocker et al. 2000). In particular, offering an explanation of why recommendations are made is deemed to influence user interaction with adaptive systems (McGuinness and Pinheiro da Silva 2004; Pu and Chen 2007). In this study, we will therefore investigate the effects of transparency by offering the users insight in the reasons why a particular recommendation has been made. To compare differences in effects between offering different types of information about a system, we will also investigate the effects of offering the users insight in how sure the system is of a particular recommendation.

This leads to three between-subject conditions to manipulate transparency:

- A non-transparent condition (‘non’): no transparency feature was offered (Fig. 1).
- A transparent (‘why’) condition: below each thumbnail of a recommended artwork a ‘why?’ link was shown (Fig. 2). This link opened a pop-up window listing those properties the recommended artwork had in common with artworks the user had previously rated positively (Fig. 3).
- Confidence rating (‘sure’) condition: below each recommended artwork, a percentage was shown which indicated the confidence the system had that the recommendation was interesting to the user (Fig. 4).

The non-transparent condition still showed pictures of the artworks the user had ‘liked’ and ‘disliked’. This communicated to the user that the system remembered the user’s preferences and did indeed tailor itself to the user. It did not show the reasons why recommendations were made.

The transparent (‘why’) condition was designed to make the criteria the system uses to recommend artworks more transparent to participants. It did so by showing the criteria on which the system had based its recommendation. Note that this is different from Wang and Benbasat’s (2007) ‘why’ condition which intended to show the recommender’s goodwill to the user’s interests.

The ‘sure’ condition’s confidence rating provided participants with information on the system’s predicted correctness of the recommendation; the system’s confidence that the user will be interested in the artwork. The sure condition was designed to

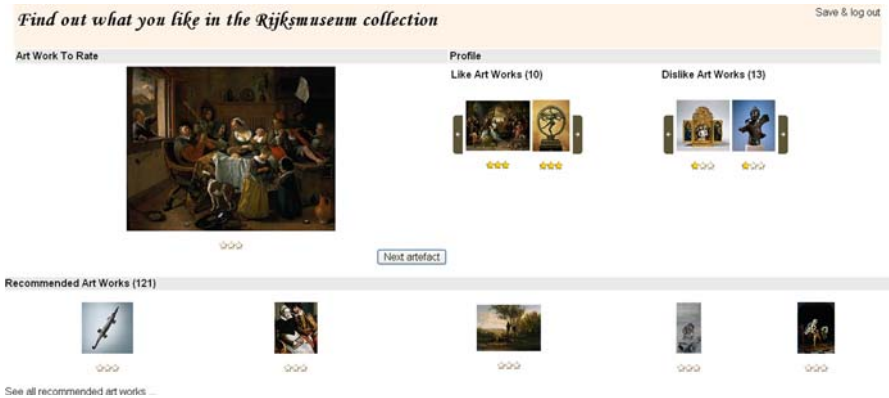


Fig. 1 Non-transparent condition version of the CHIP system

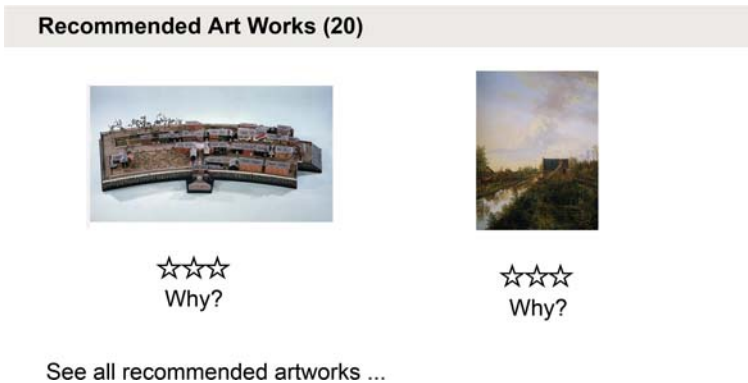


Fig. 2 Detail of the transparent ('why') version of the CHIP system, with 'why' link (replicated for printing purposes)

provide some information on system functioning, but to not provide any extra information on how the recommendations were made. The confidence ratings were intended to communicate that a level of uncertainty is involved in generating recommendations. Because participants could compare the confidence rating to their actual interest in the recommendation, this version provided users with information on recommendation competence, rather than explanation of the system's decision making.

Only the transparency features differed between versions, all other parts of the interface were the same in all of the conditions. In the top left corner, the same set of artworks was presented to the participant in the same order. These artworks could be rated by the user. The ratings were used to build the user profile. All artworks could be rated by clicking one of the three stars below the artwork. One star rating indicated little user interest; a three star rating indicated high interest. By clicking the 'next artefact' button a new artwork was shown to the user. After several ratings, recommended artworks appeared at the bottom area of the screen. The first five were shown directly; the full set of recommended artworks could be viewed by clicking

Why is “*The Tailor’s Workshop*” recommended to you?

Close

Because it has the following themes in common with artworks that you like:

- Everyday Life
- Clothes
- Occupations
- Genre painting

Fig. 3 The ‘why’ feature, pop-up window that is shown when ‘why’ link is clicked by the user (replicated for printing purposes)

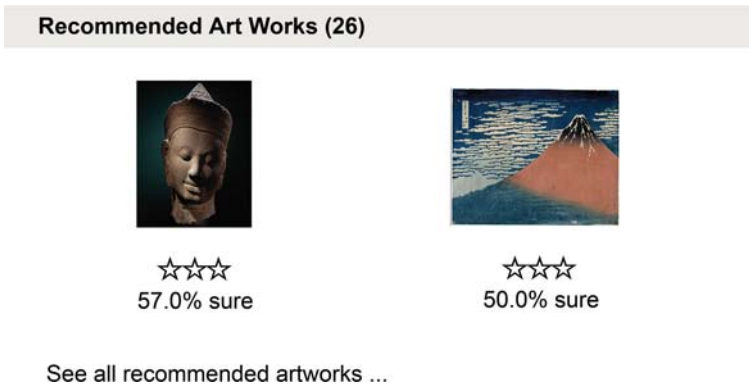


Fig. 4 Detail of the confidence rating (‘sure’) version of the CHIP system (replicated for printing purposes)

the “see all recommended artworks” button. In the transparent (‘why’) condition the “why” link was shown below each recommendation. In the confidence rating (‘sure’) condition the confidence rating was shown in the same place. In the non-transparent condition, nothing was shown below the recommendations apart from rating stars. In all conditions, the recommended artworks could be rated as well.

Users could get more information about an artwork by clicking its picture. A pop-up window was then opened showing the artwork’s title and a description. The artworks the user rated as interesting or not interesting were displayed in two sections in the upper right part of the screen. The “Like Art Works” section displayed the artworks the users rated most positive. The “Dislike Art Works” section displayed the artworks the user rated as uninteresting.

3.3 Measures

In this section, we start with an overview of the measures of the experiment. The next section will describe the procedures used in the study. Table 1 summarises the measures used in this study and provides example questions. Table 3 shows the final scales used in quantitative analysis. Questionnaire items were all seven-point Likert-type scale questions.

Table 1 Measures used in the study and example questions

Measure	Measured in:
Perceived transparency of the system	<i>Questionnaire</i> , e.g. "I understand what the system bases its recommendations on."
Actual understanding of the system	<i>Interview</i> , e.g. "Could you please tell me how the system works? It is fine if your explanation turns out not to be accurate. This question is not to test you."
Perceived competence	<i>Questionnaire</i> , e.g. "I think that the artworks that the system recommends correspond to my art interests." <i>Interview</i> , e.g. "Did the recommendations match your interest? In what aspects?"
Intent to use the system	<i>Questionnaire</i> , e.g. "The next time I am looking for a recommendation for an artwork I would like to use this system." <i>Interview</i> , e.g. "Would you use the system again for the same type of task (finding artworks for a talk)? Why?"
Acceptance of system	<i>Interview</i> , acceptance scenario
Acceptance of recommendations	Number of recommendations chosen
Trust	<i>Questionnaire</i> , e.g. "I trust the system.", "I can depend on the system." <i>Interview</i> , e.g. "To what extent do you trust the system? Why?"
Perceived usefulness and ease of use 'Sure'/ 'Why' feature ('Sure'/'Why' conditions only)	<i>Questionnaire</i> , e.g. "I found the 'why' explanations for recommendations useful." <i>Interview</i> , e.g. "How useful do you think [the 'why' feature] was?"
Perceived need for explanations	<i>Questionnaire</i> , e.g. "I think that the system should give an explanation why it recommended an art work to me." <i>Interview</i> , e.g. "Did you want the system to explain more about how it recommends art works? If so, what would you like to know?"

3.3.1 Manipulation checks

In order to check whether the transparency manipulation was successful, three seven-point Likert-type scale questions were included to measure the perceived transparency of the system in each condition with questions such as "I understand why the system recommended the artworks it did". Actual understanding was measured by asking participants how they thought the system worked and on what criteria it based its recommendations in the interview. Open questions included: "Could you please tell me your thoughts about how the system works?" and "On what criteria do you think the system bases its recommendations?" Participants' responses were compared to the way the system actually worked and scored accordingly (see the results section for a description of this comparison).

3.3.2 *Dependent variables*

Acceptance: Acceptance was measured twofold: acceptance of the system and acceptance of the recommendations. In the experiment, participants were asked to select six artworks they found interesting. Acceptance of the system was measured right after the participant had finished selecting the six artworks of interest with a scenario question: “The artworks in the current system are only 700 of a collection of 1400 pieces. Suppose you would have a chance to get a limited edition print of one of the artworks of the other half of the collection for free. You do not know what artworks are in this collection. You can choose only one. You are given two options: You would either be able to choose one artwork by going through a catalogue manually, with one artwork per page, and you only have 2 minutes to do so. The other option is that the system will take what it has learned from you in this session and will recommend three artworks from the other half of the collection. You would then be able to choose your limited edition print from these three artworks. Would you rather use the system and let it recommend you three artworks, or would you try to select one manually from a catalogue? Why?” A time limit was given for the first option, whereas a limit to the amount of paintings shown was given for the second option. This was done to ensure that participants would not think it was possible to take home a catalogue and browse it at leisure. The aim of this scenario was to create a situation of information overload, where recommenders might be a helpful alternative in order to discover whether participants intended to use the system again.

Acceptance of the recommendations was measured by the number of recommended artworks in the participant’s final selection of six artworks. For instance, a participant who had selected six artworks from the recommendations rather than from the rated artworks was thought to have a high level of acceptance of the recommendations. In addition, the level of interest in the artworks that were recommended was used as a measure of acceptance of the recommendations. A participant who scored the recommended artworks as more interesting than other artworks in the selection of six artworks was seen to accept the recommendations more.

Acceptance was further explored by open interview questions such as “Did you think the system was good at giving recommendations? Why?”, “Did the recommendations match your interest? In what aspects?” and “Would you use the system again for the same type of task (finding artworks for a talk)? Why?” Further questionnaire items that measured concepts related to acceptance were derived from items from the technology acceptance model. Participants’ attitudes towards the system were measured by questions adapted from Venkatesh et al.’s (2003) UTAUT questionnaire for technology acceptance. Three questions were included to measure intent to use the system; also adapted from Venkatesh et al. Acceptance was further explored by open interview questions such as “Would you use the system again for the same type of task (finding artworks for a talk)? Why?”.

Perceived competence of the system: Items to measure perceived usefulness and competence of the system were also adapted from Venkatesh et al. (2003); eight items were included in analysis. Perceived competence of the system and its recommendations were additionally addressed with open interview questions such as “Did you think

the system was good at giving recommendations? Why?”, “Did the recommendations match your interest? In what aspects?”.

Trust: Jian's et al. (2004) note that while a lot of authors have described the concept of trust, practical methods to measure trust appear scarce (Jian's et al. 2004). Jiang and colleagues describe how comparing results of existing scales was not feasible because of differences in the scope of the scales. Jiang et al. additionally note that these scales were predominantly based on theory and not on experimental validation. Jian's et al. (2004) have subsequently developed an empirically validated trust scale. This questionnaire scale is general in nature and needed to be adapted to the context of the system used. For instance, Jian's et al. (2004) developed a specific scale to measure trust in hybrid inspection systems for detecting faults in product manufacturing, using Jian's scale to validate their questionnaire. Jian's et al. (2004) original questionnaire was adapted and used in this study.

Twelve questionnaire Likert-type scale items were posed on trust in the system. Six out of the twelve questions used to measure trust were adapted from Jian et al. (2000). The other questions were designed by the first author and focused on measuring to what extent the participant trusted the system to be competent. Ten questions were used in analysis.

An open interview question further explored participants' trust toward the system: “To what extent do you trust the system? Why?” and “For what type of activities do you think you would trust the system?”.

Attitudes towards the ‘why’ and ‘sure’ features: Participants' attitudes toward the ‘why’ and ‘sure’ features were explored in both the questionnaire and interview items. Participants were also asked whether they thought the system should explain more about the way it worked with the scaled questionnaire item “I think that the system should give an explanation why it recommended an art work to me.” An open interview question explored this further: “Did you want the system to explain more about how it recommends art works? If so, what would you like to know?”

Two to six additional scaled questions were offered in each condition, which addressed usefulness, ease of use and understandability of the specific transparency feature offered. In the non-transparent condition two additional questions were asked on the need for transparency information, e.g. “An explanation on why an artwork has been recommended would be useful”. In the ‘sure’ and ‘why’ conditions six questionnaire scale items such as “I found the ‘% sure’ percentages for recommendations useful” and “I thought the ‘why’ explanations were clear and understandable,” were asked.

Interview questions, both open and closed, further investigated participants' opinions of how useful an offered transparency option (if applicable) was and whether they thought the system should explain recommendations. Nine interview questions were included specifically to evaluate the specific transparency feature offered in the ‘sure’ and ‘why’ conditions, with questions such as: “Did you notice the [‘sure’ / ‘why’] feature in the system” and “Did they influence how you felt about the system? In what way?”.

Demographics: Participants' demographics and other background variables that could influence acceptance of the system were measured, including age, gender, computer experience, experience with recommender systems, level of education, interest in art and knowledge of art.

Long-term interests and attitudes: Long-term perceived competence of the system (for example “I like the artworks the system recommended to me”), trust in the system (for example “I am confident in the system”) and intent to use the system again (for example “The next time I am looking for a recommendation for an artwork I would like to use this system again”) were measured in a follow-up experiment session.

3.4 Procedure

Each participant participated in one of the three conditions and took part in individual, task-oriented sessions that lasted 45 min to three hours. Each individual session with a participant consisted of three parts: (1) observation of the participant carrying out a task using the CHIP system (described in more detail below) (2) a questionnaire evaluating participants’ acceptance and trust in the system and (3) a structured interview to further explore participants’ attitudes towards the system. The study was conducted by two researchers; each observed 50% of the participants for each of the three conditions. Eight weeks after their session, participants were sent a follow-up questionnaire to evaluate participants’ longer-term attitudes towards the system. Each of these parts of the study is explained below.

3.4.1 Task observation

At the beginning of each session, the experimenter introduced the study and the CHIP system. Briefly explaining the purpose of the CHIP system, the researcher explained how to rate artworks (by clicking one of the three stars), where recommendations would appear (in the lower section of the screen), how to see the full list of recommendations (by clicking the ‘see all artworks’ link). The researcher also explained that the system would show their liked and disliked artworks, showed how to get more information about an artwork (by clicking its picture) and explained the ‘why’ or ‘sure’ features if applicable.

Participants were given the task to use the CHIP system to select six art works to develop a short presentation about their personal art interests. During this task, participants were observed by the researcher and were asked to think aloud. The sessions were audio recorded and participant behaviour was collected by screen capture software and log files. The researchers took notes paying attention to comments about the system and reactions to its recommendations, the ‘why’ and ‘sure’ features and to usability difficulties the participants encountered.

When participants finished after selecting six artworks, they were asked to paste their selection in a word document and indicate for each artwork how interesting they found the artwork on a seven-point Likert-type scale “I think this artwork is really interesting.”, ranging from “very strongly disagree” to “very strongly agree”. Observation sessions lasted from 45 minutes to 3 hours.

3.4.2 Questionnaire

After completing the task, participants were asked to fill out a questionnaire. The post-task questionnaire included questions addressing participant’s demographics, and 41

seven-point Likert-type scale questions addressing perceived usefulness and competence of the system, attitude towards the system, intent to use the system, trust in the system, and two (non-transparent condition) to six ('why' and 'sure' conditions) items measuring attitudes towards transparency information.

3.4.3 Interview

An interview concluded the session. Participants were first asked to explain for each of the six selected artworks why this artwork was interesting to them. Then the acceptance scenario was presented to the participants. The interview's fifteen questions then addressed participants' perception of competence of the system, participants' trust in the system, understanding of the system and their reasoning while rating artworks and attitudes towards transparency.

3.4.4 Follow-up questionnaire

After eight weeks, participants were sent an email with a link to an online questionnaire. The questionnaire first tested participants' satisfaction with their previous selection of six artworks. As a reminder, a picture of each of the artworks they had selected was included, and participants were asked to indicate again for each artwork how interesting they thought the artwork was on a seven-point Likert-type scale. The questionnaire then showed participants the first five recommended artworks they had received during their first session. Thirteen questions then addressed long-term perceived competence, trust and intent to use the system again.

3.5 Participants

During a period of three weeks, 82 participants were involved in the study. Twenty-two participants were excluded from analysis. Sixty participants were included in quantitative analysis: twenty-two participants were included in the non-transparent condition, nineteen participants were included in the transparent condition and another nineteen were included in the 'sure' condition. Participants were excluded from quantitative analysis when the system had not provided them with any recommendations during their session (this could happen when they rated a large number of artworks, but only rated a very limited number of artworks as interesting), if they had not noticed the recommendations in the lower section of the screen, had not used the 'why' feature in the transparent condition or if technical problems were encountered (e.g. time-outs of the application). The data from excluded participants were included in qualitative analysis to gain information on possible usability issues and issues with the system's recommendation strategies in order to develop guidelines for recommenders in the art domain.

Participants were volunteers drawn from the researchers' professional and personal networks. These participants suggested others who may also be interested in participating in the study. Because this might have led to an overrepresentation of people

Table 2 Participant characteristics

Age	Mean = 34.4 years (SD = 13.63, range 18–68)
Gender	Male $N = 31$, Female $N = 29$
Maximum level of completed education	1.67% ($N = 1$) primary education 13.3% ($N = 8$) secondary education 21.7% ($N = 13$) receiving/completed professional tertiary education 63.4% ($N = 38$) receiving/completed academic tertiary education
Previous experience with recommender systems	31.7% ($N = 19$) previous experience
Computer experience	Mean = 5.43, SD = 1.23, on a 1–7 scale
Knowledge of art	Mean = 3.55, SD = 1.67, on a 1–7 scale, 13.3% high level
Interest in art	Art in general: mean = 5.10, SD = 1.57, on a 1–7 scale Art in Rijksmuseum: mean = 4.45, SD = 1.44, on a 1–7 scale

who are interested in technology or art, computer science and art background was measured in the experiment. Participants were relatively well educated and experienced in using computers. None were experts on recommender systems. Both art experts and non-experts were involved in the study. Table 2 offers an overview of participant characteristics.

The demographics age, gender, computer experience and interest in art were not found to have a significant influence on the results of this study. However, female participants reported significantly less experience in using computers than male participants (Mann–Whitney $U = 270.500$, $p = .006$, $N_m = 31$, $N_f = 29$). Computer experience however was not found to be correlated with other variables.

4 Results

This results section is structured as follows: first the effects of transparency on perceived understanding, perceived competence, acceptance and trust are discussed. The effects of offering confidence percentages as in the confidence rating ('sure') condition are discussed afterwards. Possible changes in long-term attitudes towards the system are addressed as well. The main correlations between the variables are discussed in the last subsection. Possible explanations and implications of the results described in this section are discussed in the subsequent discussion (Sect. 5). Table 3 shows the final scales used for analysis, reliability of the scales and means scores of our main variables of interest. Cronbach's Alpha was used to determine scale reliability. Questions that decreased scale reliability were excluded. To test our hypotheses, we conducted Mann–Whitney analyses as an alternative to t -tests as the data did not meet parametric assumptions.

Table 3 Final scales and variables included in overall analysis, including Cronbach's alpha for the final scale. Questionnaire items were seven-point Likert-type scale questions, with scale ranging from 1 ('very strongly disagree') to 7 ('very strongly agree')

Perceived transparency of the system

Cronbach's Alpha = .7433

Mean = 4.52, SD = 1.19, range: 2.0–6.5

I understand why the system recommended the artworks it did.

I understand what the system bases its recommendations on

Perceived competence

Cronbach's Alpha = .914

Mean = 4.0687, SD = 1.14920, range: 1.63–6.50

I think that the system's criteria in choosing recommendations for me are similar to my own criteria.

I like the artworks the system recommended to me.

I think the system should use other criteria for recommending artworks to me than it uses now. (inverted for analysis)

The artworks that the system recommended really interest me.

I think that the artworks that the system recommends correspond to my art interests.

I think the system does not understand why I like certain artworks I rated as interesting. (inverted for analysis)

I think the system does a good job adapting to what I tell it to be interesting artworks.

The system correctly adapts its recommendations on the basis of my ratings.

Intent to use the system

Cronbach's Alpha = .914

Mean = 4.3833, SD = 1.35828, range: 1–6.67

I would rather choose the 6 artworks by hand from the collection of artworks than use the system if I would have to perform this task again. (inverted for analysis)

I would like to use the system again for similar tasks.

The next time I am looking for a recommendation for an artwork I would like to use this system.

Acceptance of Recommendations

Mean = 2.04, SD = 1.86, range: 0–6

Number of recommendations that the participant included in their final selection of 6 artworks.

Acceptance of the System

Participants choosing system selection = 32, Participants choosing manual selection = 28,

Total $N = 60$

Scenario measuring participant's willingness to delegate task to system.

Trust

Cronbach's Alpha = .901

Mean = 4.2, SD = 1.06, range: 1.8–6.2

I am confident in the system

I trust the system not to recommend artworks that are not interesting to me.

The system is deceptive. (inverted for analysis)

I trust the system to recommend me all artworks that are of interest to me.

The system is reliable.

Using this system for these tasks is risky. (inverted for analysis)

I trust the system.

Using this system is risky. (inverted for analysis)

I can depend on the system.

I trust the recommendations of the system to match my own selection.

4.1 Effects of transparency on perceived and actual understanding

As expected, participants did find the transparent ('why') condition more transparent than the non-transparent condition (Mann–Whitney $U = 136.500$, $p(1 - \text{tailed}) = .025$, N 'non' = 22, N 'why' = 19, Mean'non' = 4.4, SD 'non' = 1.1, Mdn 'non' = 5.0, Mean'why' = 4.9, SD 'why' = 1.4, Mdn 'why' = 5.0). Participants in the confidence rating ('sure') condition did not think they understood the system better than those in the non-transparent condition (Mann–Whitney $U = 195.500$, $p(2 - \text{tailed}) = .720$, N 'sure' = 19, Mean'sure' = 4.3, SD 'sure' = 1.1, Mdn 'sure' = 4.0). This finding shows that offering explanations for the systems recommendations indeed causes participants to feel they understand the system better, while offering a confidence percentage does not. However, these results relate to participants' self-reported, perceived understanding.

Whether offering explanations also increased actual understanding of the system was investigated by content analysis of participants' answers to the interview questions about understanding. Because of resource constraints and its time-consuming nature, this qualitative analysis was limited to the participants of one of the researchers. Only answers of participants who participated in the non-transparent or 'why' condition ($N = 21$) were analysed. Comments made by these participants concerning composition of the user profile, the rating process and recommendation process were individually rated by three researchers for accuracy. Participants' answers to these questions were divided in atomic statements; duplicate statements in a single participant's answers were removed. A participant's 'understanding score' was then calculated by dividing the total number of accurate comments by the total of correct and incorrect comments. Three coders coded statements to verify the coding scheme's reliability. Inter-coder reliability was analysed by computation of the Intraclass Correlation Coefficient. The ICC can range between 0 (indicating no interrater agreement) and 1 (indicating total interrater agreement). In this study, the intraclass correlation coefficient produced a value of 0.808, which can be considered to be very good [Nückles et al. \(2006\)](#). Having obtained an acceptable interrater reliability, subsequent analysis was carried out on the data of the main coder only. A similar procedure is followed in [Kurasaki \(2000\)](#).

Interesting is that for the participants whose understanding was analysed, there was no significant correlation between how well participants thought they understood the system and their actual understanding (Spearman's $\rho = .005$, $p(1 - \text{tailed}) = .491$, $N = 21$). This illustrates the necessity of measuring participants' actual understanding as well as participants' perceived understanding; not all participants appear capable of assessing their own level of understanding of a system.

A significant difference was found between actual understanding of subjects in the non-transparent condition ($Mdn = 7.14$) and actual understanding of subjects in the transparent ('why') condition ($Mdn = 9.50$) on how the system works, Mann–Whitney $U = 20.000$, $p(1 - \text{tailed}) = .0065$. Participants in the transparent 'why' condition were found to have a better understanding of how the system came to its recommendations; this confirms our manipulation of transparency in the 'why' condition was successful.

4.2 Effects of transparency on perceived competence

The data did not support our hypothesis that a system with a more transparent decision-making process will be perceived as more competent. Scores on perceived competence of the systems were compared using the Mann–Whitney test between the non-transparent condition and the transparent ('why') condition. No significant difference was found between the conditions (Mann–Whitney $U = 190.000$, $p(1 - \text{tailed}) = .310$, $N\text{'non'} = 22$, $N\text{'why'} = 19$, $\text{Mdn}\text{'non'} = 4.0$, $\text{Mdn}\text{'why'} = 4.5$). It appears that transparency does not necessarily involve an increase in the perceived competence of a system.

4.3 Effects of transparency on acceptance

To test our hypothesis that transparency would increase system acceptance, both acceptance of the system and acceptance of the system's recommendations were compared for the non-transparent and transparent ('why') condition. The final selection of artworks for participants in the transparent ('why') condition contained significantly more recommendations than in the non-transparent condition (Mann–Whitney $U = 125.500$, $p(1 - \text{tailed}) = .013$, $N\text{'non'} = 22$, $N\text{'why'} = 19$, $\text{Mdn}\text{'non'} = 1.0$, $\text{Mdn}\text{'why'} = 2.0$). Acceptance of the recommendations was indeed higher in the transparent condition.

However, when investigating acceptance of the system using the scenario question that measured participants' willingness to delegate a task to system, no significant difference was found. Participants in the transparent condition were not more willing to delegate the task of selecting artworks to the recommender than the participants in the non-transparent condition. In the non-transparent version 11 of 22 participants chose the system, while in the 'why' condition 10 of the 19 participants chose the system; not a significant difference. Participants' intent to use the system were compared using the Mann–Whitney test (Mann–Whitney $U = 194.500$, $p(1 - \text{tailed}) = .352$, $N\text{'non'} = 22$, $N\text{'why'} = 19$, $\text{Mdn}\text{'non'} = 4.5$, $\text{Mdn}\text{'why'} = 5.0$). No significant difference was found.

Our hypothesis that a more transparent decision making process would increase participants' acceptance of the system can only be partially accepted. The data suggests that it is important to differentiate between acceptance of recommendations and acceptance of the recommender system. The data of this study indicate that transparency does indeed influence acceptance of the system's recommendations, but that the transparency feature offered here does not influence acceptance or adoption of the system itself.

4.4 Effects of transparency on trust in the recommender system

Contrary to our hypothesis, participants in the transparent condition did not trust the system more. Scores on trust between participants in the non-transparent condition and the transparent ('why') condition were compared using the Mann–Whitney test (Mann–Whitney $U = 180.000$, $p(1 - \text{tailed}) = .224$, $N\text{'non'} = 22$, $N\text{'why'} = 19$,

Mdn 'non' = 4.2, Mdn 'why' = 4.7). No significant difference was found on the scores on trust in the system between the conditions.

4.5 Effects of certainty ratings: the 'sure' condition

Offering different types of information about a system was expected to have different effects on attitudes towards a system. The 'sure' condition provided certainty ratings in a numerical format. It did not provide insight in the system's reasoning process and the reasons behind the recommendations. Participants in the confidence rating ('sure') condition did not perceive the system to be more transparent. This was expected, as the 'sure' feature did not offer participants insight in the criteria the system uses to choose recommendations. It was not expected to significantly increase understanding of the system's reasoning process. Scores on perceived competence of the systems were also compared between the non-transparent condition and the 'sure' condition (Mann–Whitney $U = 170.000$, $p(2 - \text{tailed}) = .307$, N 'non' = 22, N 'sure' = 19, Mdn 'non' = 4.0, Mdn 'sure' = 3.9) as were scores on intent to use (Mann–Whitney, $U = 165.000$, $p(2 - \text{tailed}) = .247$, N 'non' = 22, N 'sure' = 19, Mdn 'non' = 4.5, Mdn 'sure' = 4.0). No significant differences were found.

No significant differences were found either on acceptance of the system and its recommendations between the non-transparent and 'sure' condition. The final selection of artworks for participants in the 'sure' condition did not contain more recommendations than in the non-transparent condition (Mann–Whitney $U = 168.000$, $p(2 - \text{tailed}) = .271$, N 'non' = 22, N 'sure' = 19, Mdn 'non' = 1.0, Mdn 'sure' = 2.0). In the acceptance scenario, participants in the 'sure' condition were not more willing to delegate the task of selecting artworks to the recommender than the participants in the non-transparent condition. Participants also did not trust it more than those in the non-transparent condition did (Mann–Whitney $U = 187.500$, $p(2 - \text{tailed}) = .573$, N 'non' = 22, N 'sure' = 19, Mdn 'non' = 4.2, Mdn 'sure' = 3.8).

It appears that not all information that can be offered to the user about system decisions leads to improved transparency, acceptance and trust. Participants additionally reported to find the certainty percentage slightly not useful on a 1–7 Likert-type scale question (Mean = 3.70, $N = 19$, $SD = 1.45$), making positive effects on trust and acceptance of the feature less likely.

4.6 Long-term attitudes towards the system and its recommendations

Fifty-three of the 60 participants responded to the follow-up email sent eight weeks after their first session and filled out an online follow-up questionnaire. The questionnaire addressed long-term perceived competence, trust and intent to use the system again. Participants were presented with the artworks they had chosen as their favourites and the first five of the system's recommendations. Participants rated the artworks they had chosen as their favourites in the first sessions as interesting in the follow-up questionnaire as well; their art preferences and attitudes towards the system were more or less constant. Results on perceived competence, trust and intent were the same in the first session and the follow-up questions. Kruskal–Wallis tests were used to test for

differences between the conditions on these scores and whether scores had increased or decreased more in either of the conditions. No significant differences were found. The constructs as measured in the first session can be taken as representative of longer-term opinions of the participants. A possible implication might be that user preferences in the e-culture domain might be relatively stable and profiles might be valid during a longer period after first training a system.

4.7 Correlations for perceived understanding

Based on our hypotheses we expected perceived understanding to correlate with perceived competence, trust and acceptance. Spearman's rho was used to calculate correlations between perceived understanding and the main variables (Fig. 5). Perceived understanding did correlate with trust and system acceptance. Perceived understanding of the system (Spearman's rho = .517, $p(2 - \text{tailed}) = .000$, $N = 60$), and perceived competence of the system (rho = .663, $p(2 - \text{tailed}) = .000$, $N = 60$) also were correlated to participants' reported intent to use the system. These findings confirm the expected relationships between variables; they suggest that perceived understanding relates to system acceptance, which is the user's choice to delegate a task to the system. Unexpectedly however, perceived understanding did not directly correlate with acceptance of the recommendations, nor did acceptance of recommendations correlate with acceptance of the system. This last finding illustrates the need to distinguish between acceptance of a user-adaptive system as a whole and its recommendations, suggestions or decisions.

Trust is correlated with intent to use (rho = .805, $p(1 - \text{tailed}) = .000$, $N = 60$), the decision whether to use the system or catalogue (rho = .453, $p(1 - \text{tailed}) = .000$, $N = 60$) and to the acceptance of recommendations (rho = .363, $p(1 - \text{tailed}) = .002$, $N = 60$). We conclude that trust is related to the decision whether to use the system.

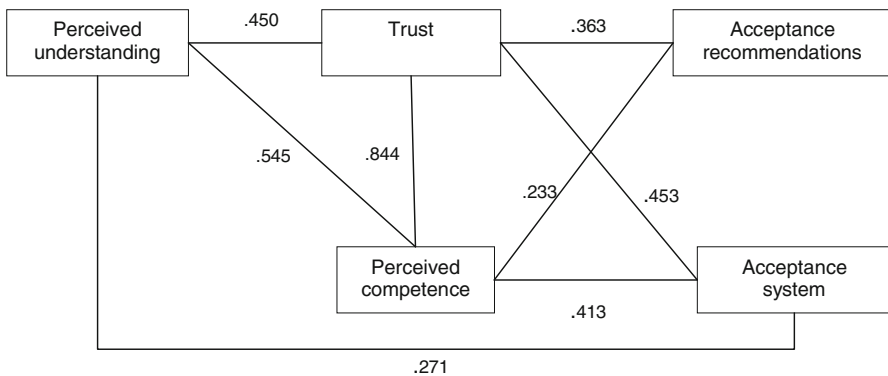


Fig. 5 Main significant correlations, Spearman's rho, $p(1 - \text{tailed}) < .05$

5 Discussion

This section offers a discussion of the results described in the section above. We indicate a number of limitations to the study and features of the conditions that might have influenced results. Participants' responses and observations from the participant sessions are discussed to explore further our quantitative results. Implications of this study and guidelines for user-adaptive systems in a cultural heritage context are discussed as well.

5.1 The effects of transparency on understanding

The manipulation of transparency was effective in our study. The results of this study show that the transparent version of the recommender system was indeed better understood. Participants in the transparent condition reported higher perceived understanding and were also found to have more actual understanding of the recommendation criteria. When asked about their understanding of the system in the interview, participants with the lowest degrees of understanding offered incorrect views about how the system worked. For example, one participant in the non-transparent condition reported that the artworks were divided into categories, which is incorrect since there are multiple features or properties associated with artworks. Another participant in the non-transparent condition could not comment anything about how the system works. Finally, another participant in the non-transparent condition thought that the system used collaborative filtering techniques to provide recommendations, when in fact the system relies on content-based techniques. Such misunderstandings can be countered by offering users explanations on how a system selects the recommended items. Participants with the highest degrees of understanding gave more detailed descriptions of artwork features or properties attached to artworks, and reported that the system provides recommendations based on similarities between artworks based on their properties.

Understanding of the system led to some emergent behaviour for some of our participants. Subjects in the transparent 'why' condition also reported 'filtering behaviour' in their interaction with the system; these subjects indicated that they were actively adapting their rating behaviour to make the system provide different and better recommendations. For example, some participants noted that when they rated Asian art as interesting, the system would provide related themes in explanations. They would however not only want Asian art recommendations and therefore would rate some Asian art as less interesting than they would have otherwise. This illustrates how transparency and understanding can influence user feedback to a system. It is important to make a system more understandable to increase acceptance of its recommendations, but also to ensure that user feedback increases perceived and actual performance of the system. Making a system more transparent without offering users a direct way to correct a system's mistakes will be frustrating and is likely to affect attitudes towards the system negatively.

Which aspects of a system's decision making process are better understood when users interact with a more transparent system is likely to be dependent on the specific type of transparency offered. In all of the conditions, including the non-transparent

condition, the systems showed pictures of the artworks that the user had rated positively and negatively. One could argue that this communicated to the user that the system ‘knew’ of the user’s likes and dislikes, and thus provided information on how the recommendations were made. As such, all conditions gave some information concerning the fact that the system used positive and negative user ratings. The transparent condition however was still better understood and led to higher acceptance of the recommendations. This suggests that transparency can indeed improve user interaction with user-adaptive systems.

5.2 The effects of transparency on perceived competence

We expected that a system with a more transparent decision making process would be perceived as more competent, but this was not the case in our study. A number of reasons can be offered for this finding. Participants in the transparent condition did not always agree with the criteria included in the explanation why an artwork had been recommended. One participant for example, who was an art expert, disagreed with some of the labelling of artworks. He thought the criteria and explanations were too simplistic. This is in line with previous literature where users who are familiar with the content are expected to be more critical of the system and perceive it as less dependable (Fogg and Tseng 1999). It could be that the transparent condition made discrepancies between the user’s decision-making process and that of the system more obvious. Even when a participant liked a recommendation, the criteria on the basis of which the system recommended the artwork and the reasons why the participant had chosen the artwork could still be different. However, scores on perceived competence were not different in the transparent condition, indicating that this was not a deciding issue in this study. Explanations could also help convince users of a system’s competence, if an explanation itself is acceptable to them. An extensive discussion of the perceived competence of the recommender, its relation to actual competence and effects of transparency on this relation can be found in Cramer et al. (2008).

5.3 The effects of transparency on acceptance

A more transparent decision making process was expected to increase users’ acceptance of the system. This could only be partially confirmed in our study; only recommendations were accepted more. Why are users more accepting of a system’s recommendations if they are more transparent, but is a more transparent system not necessarily adopted more? Explaining why a particular recommendation had been made, might have helped acceptance of that recommendation. It might have helped participants accept individual artworks or increased their liking for artworks they found interesting, but did not convince them to adopt the system more. In response to the interviews’ acceptance scenario, participants expressed a preference for selecting artworks by hand. The decision to use or not use the system appeared to depend more on whether an acceptable artwork could be found by browsing the catalogue within the time available or that an efficient recommender tool would be needed. If there is no perceived information overload, a recommender system appears unnecessary.

System competence and transparency appeared to be secondary to these considerations in overall system acceptance. There is a possibility that explanations make it easier to critique a system and could in some cases have negatively influenced trust and acceptance. However, this appears not to be the main cause of the absence of the expected positive effect of transparency in this study; no significant differences were found between conditions in perceived competence of the system.

It is possible that in other situations, transparency of recommendation can help acceptance of a system as well, but whether the system will actually be used depends on a large number of other factors, such as perceived competence of the system and whether the user actually feels the need to use a system. Indeed, perceived understanding of the system and perceived competence of the system were correlated to participants' reported intent to use the system. This is consistent with existing technology acceptance models (Venkatesh et al. 2003).

5.4 The effects of transparency on user trust

Trust was positively correlated with perceived understanding of the recommender system. Transparency aimed at increasing understanding of a system can thus play a role in building trust. However, the hypothesis that a system with a more transparent decision-making process will be trusted more could not be confirmed in this study. The transparent version of the system used in this study was not trusted more. Some of the possible reasons behind the absence of effects of transparency on acceptance of the system (in contrast to acceptance of recommendations, where a positive effect was found), might also play a role in this absence of effects on trust.

A number of issues specific to the cultural heritage domain of the application could play a role in these results. Trust in the system appears to be mainly based on the perceived competence of the system. It is possible that users could more easily identify limitations of the system in the transparent condition. This could be related to the relatively low scores on perceived competence of the system, (mean = 4.07 on a 1–7 scale, SD = 1.15, range: 1.63–6.50, only slightly above neutral). Participants noted that the system could currently not address all of their art interests. Only concepts defined by experts could be used for the content-based recommendations. Many of the aesthetic and affective interests that play a role in the likes or dislikes of participants in art (e.g. their personal interests, or emotional reactions) did not appear to be addressed. Participants were able to identify unsuitable criteria used for recommendations in the transparent version. Possibly, the perceived limitations of the system might have cancelled out any positive effects of transparency on trust.

Another issue specific to the context of cultural heritage and art is that recommendations involved static visual artworks, such as paintings, or photographs of historical objects and visual objects such as statues. Due to the nature of these recommendations, it was possible to evaluate quality of the recommendations instantly. Participants could immediately see the pictures of the recommended artworks and decide whether they liked them or not. This is in contrast with, for example, movie recommenders, where users first need to watch a recommended movie before they know whether the system's recommendations are any good. The risk to the user in following up on a recommen-

dation is low as the recommendation itself and the content of the recommended item were largely the same. Additionally, risks of receiving less interesting recommendations are relatively low in an e-culture context. Consequences of seeing less interesting artworks instead of the ones you would really be interested in, are not very serious.

Perhaps the concept of trust plays less of a role in interaction with user-adaptive recommenders, especially in cases where direct evaluation of the correctness of recommendations is possible and risks associated with system mistakes are small. It can also be argued that when the intentions of a system are clear and match the users' goal, the concept of trust is only partially applicable as rather only competence-related aspects of trust are of interest. In the cultural heritage domain, trust might for example be more relevant in contexts where the correctness of information provided about the artworks might be unclear. Trust in perceived competence appeared to be of main importance in this study.

Self-reporting of trust appeared somewhat problematic in this study's interviews. Interesting is that participants did report trusting the system, but that this was not actually translated into willingness to let the system choose artworks for them. They still would rather choose themselves. It could be argued that focus should be on the willingness to delegate to a system, instead of trying to measure trust in such situations. While trust appears to be a popular research concept in human-computer interaction, currently available methods to measure trust appear to lack standardisation. Widely accepted, standard scales that are applicable to interaction with user-adaptive systems are not available (yet). While the scale we used can for example be adapted and re-used in other studies, further research is necessary addressing this issue.

5.5 Effects of certainty ratings on interaction

The study findings show that offering explanations had a significant effect on user understanding and acceptance of the recommendations. The transparency feature that showed certainty ratings however, did not have such an effect. The interviews offered some insight into why the confidence rating ('sure') condition did not impact trust and acceptance. Seven participants commented on the discrepancy of their interest in artworks and the percentages given by the system. One participant reported liking an artwork labelled as 8% sure, while disliking one the system was 27% sure about. Another said he understood why a particular artwork had been labelled as 82% sure and that the percentage fit the ratings he had given, but that he did not really like the artwork itself.

Generally, percentages were under 50% during most of the participants' session as the system was building training data to become 'more sure' of its recommendations. One participant remarked that this made it appear as if the system was not sure at all of what the participant liked. On the other hand, another participant appreciated knowing that the system was not sure of its recommendations and that recommendations were not "pushed" by the system. The knowledge that the system was unsure about its recommendations made some participants less inclined to trust the system.

The certainty ratings were not always understood. Typically, in rule-based systems, each argument contributing to a conclusion is given a certainty factor. Together these factors build up to the overall certainty of a conclusion. One single certainty percentage, without really understanding in what way the system calculates the rating may

leave the user to wonder what the rating really represents. Percentages appear difficult to interpret; any indicators of system performance or (un)certainty should be provided in a form that is meaningful (e.g. very sure, or unsure, instead of 70% or 10%). More general and widely adopted retrieval systems such as Google also sort results using numeric relevance measures, without showing the numbers themselves. Lee and See (2004) also note that abstract data, such as content-free percentages, lead to different risk assessments than when specific instances and salient images are used. More concrete data might be just as, or more, powerful in risk assessment and their effect on trust. McNee et al. (2003) have shown that simple certainty ratings can indeed invoke more appropriate usage decisions.

Unforeseen effects of certainty ratings on feedback behaviour have to be considered as well. Four participants explicitly reported that the percentages influenced their behaviour. They said that the percentages invited them to rate more artworks or change ratings to make the system surer about recommendations they liked. One decided to always rate all the recommendations he disliked so that the system would be more than 50% sure, perceiving 50% as a cut-off rate for interest. Another aimed to get the certainty percentages for liked artworks higher. One other participant stated that the sure percentages sometimes decreased after giving more ratings. The participant found that this reduced motivation for training the system further.

These findings indicate that there is not a one-size-fits-all solution in making user-adaptive systems more transparent. Future research is needed to fully understand what combination of transparency features will optimise interaction with user-adaptive systems.

5.6 Implications for designing user-adaptive systems in the cultural heritage domain

During observation of the 82 participants while they interacted with the art recommender, interesting observations were made that can inform the development of other user-adaptive systems that try to personalise cultural heritage explorations.

5.6.1 *System criteria need to match user and contextual criteria*

During the interview, participants were asked the reasons why they liked the six artworks they had chosen as their favourites. Often, participants would not mention artists, themes or other content-related features the system used to recommend artworks. Instead, they would refer to more personal or emotional criteria for liking an artwork. They would quote the depiction of dynamic movement in a picture or its colours as reasons for liking it. Participants mentioned personal experiences as well, providing for example information about family history or family anecdotes that were related to a particular artwork. One of our participants explained she did not particularly like a specific artwork, but did find it very interesting. Another commented she did not find an artwork interesting at first, but that reading its description had sparked her interest. This illustrates the complexity of user feedback indicating which artworks are interesting to them. The features that are important in the cultural heritage domain, specifically in the context of art, can be difficult for adaptive technology to deduce.

Generating a personalised tour in a museum might require different information than generating an educational presentation, as might finding a single artwork the user would like. It could be useful to include user defined themes (e.g. tags) and collaborative-based recommendation to address shared interests or criteria not described by expert annotators to include e.g. more personal or emotional criteria for liking an artwork.

5.6.2 Offering recommendations at the right time during the interaction process

In this study, a number of participants had to be excluded from quantitative analysis because the system did not provide them with any recommendations. In a few cases, this was not because they did not like any of the art in the system's collection, but because the system did not accommodate the users' rating style. These participants could be classified as 'tough critics'; they generally rated artworks with a neutral rating and did not rate many artworks positively. Consequently, the system did not have enough properties to offer recommendations. In this study, some of these participants expressed offence at more or less 'being told' by the system that they had no interest in art at all. This illustrates the need to carefully consider when to introduce recommendations in the interaction process.

5.6.3 Adapt gradually, clearly and offer restore options

In this study, a negative rating sometimes caused the system to immediately eliminate all artworks from its recommendations that had been annotated with the same themes. Some participants were frustrated by these changes in the set of recommended artworks and became weary of correcting the system, fearing that they would eliminate interesting artworks as well without having a way to retrieve them again. Gradual adaptation and ways to restore eliminated options are necessary.

5.6.4 Consider interface effects on users' mental model of the system

Some participants in this study suspected more adaptation than was actually implemented. The system used in this study did not adjust the 'training set' of the artworks the user could rate (seen in the upper left corner of the screen, Fig. 1). However, some participants commented on how well the system adjusted the training set to their preferences, even though the order of artworks was fixed for all participants. Some of these participants did not even notice the systems' actual recommendations. This observation suggests that the promise of adaptation can easily lead to an incorrect mental model of the system; interface design needs to take this into account and clearly identify the adaptive elements of the interface.

5.6.5 Understandable and meaningful explanations

As [Herlocker et al. \(2000\)](#) found, explanations that are too complex might actually negatively affect acceptance of a system. In this study some participants were confused by themes in explanations they could not relate to the recommended artwork.

For example, some participants encountered the theme ‘militia paintings’. Not all participants understood what a ‘militia painting’ was and why it would be interesting to them. Such expert annotations of cultural heritage information are often difficult to understand for the laymen users. One of the expert users however, found the same term too simplistic. These anecdotes argue for not only profiling users to offer recommendations but also to offer user-adapted explanations that are meaningful to the specific user. Explanations have to be tailored to the goal of a system as well. When generating a personalised tour, they may for example need to focus on relationships between artworks instead of the properties of a single artwork. Additionally, explanations might be used to seize the opportunity to educate the user about the styles and techniques he or she appears to find interesting.

5.6.6 Consider potential misconceptions

Transparency features such as explanations cannot be developed in a vacuum from the rest of an interface. In the ‘why’ condition users’ mental models of the recommendation process were not only influenced by the explanations. The system used in this experiment provided both an explanation and a regular description of the artwork. A limited number of participants thought the system used the descriptions available for each artwork to recommend artworks. Explanations thus might need to counter misconceptions users have formed while interacting with the rest of the system’s interface. Unexpected effects of the explanation itself need to be considered as well. For example in our study, a limited number of participants in the ‘why’ condition indicated when seeing an explanation for the first time, that the system based its recommendations on too few criteria. It appeared that when they saw that an explanation for recommendation only listed one or two properties it “had in common with other artworks they liked”, they thought the artwork itself had only been annotated with one or two properties in total, while in reality the artworks had been annotated with many more criteria.

5.6.7 User freedom and control

As [Jameson \(2003\)](#) points out, users have a need for control that is in conflict with automatic user-adaptivity. If a system’s criteria are transparent, ways should be provided to correct them if they appear unsuitable to the user. Knowing that a system’s criteria are wrong, but not having any way to correct them could frustrate users and negatively affect potential positive effects of transparency of system acceptance and trust. More control over the system’s profile might also be useful to accommodate users that already have some idea of the type of recommendations they are looking for. While users in the cultural heritage domain might be interested in learning about new artworks, our participants appeared to have an idea of the art they find interesting if they are looking for an entertaining personalised experience. These users should not only be provided with a training set, but also be provided with (understandable) options of explicitly including or excluding categories of art in the system’s collection. Issues such as posed by [Waern \(2004\)](#) who discusses how users cannot always improve on their own user profile, should be taken into account.

5.7 Limitations of this study

There are a number of limitations to the results of this study regarding the effects of transparency on trust and acceptance. First of all, only two features that provided participants with some insight in how the system worked have been implemented and evaluated in this study. Although careful deliberation preceded the choice for these specific features, the implementations did not turn out to be optimal. The study discovered several issues surrounding the implementations and the features' design were encountered during participants' interaction with them. Sections 5.5 and 5.6 discuss such limitations and give a number of examples. Using other types of implementations of transparency features, with more direct control to correct system mistakes could have yielded different results.

In addition, there are limitations to generalisation of this study's findings to other domains. The visual nature of the recommendations facilitated direct evaluation of the recommendation results. This is not the case for every domain, e.g. in case of movie recommendations, direct evaluation of the movie is not possible; you have to see the movie first. Participants additionally did not have to deal with real, high-impact consequences when the system would have made a mistake. Such risks are relatively low in the art domain. In more risky situations where recommendations of filtering results cannot be immediately evaluated, trust and transparency might play a greater role. Additionally, during this experiment, participants interacted with the system in one session. The effect of transparency might be different when a system is used repeatedly during a longer period of time. Users might for example be more interested in transparency features during their first encounters with a system and later only be interested in explanations when a system presents them with unexpected results. Follow-up research is needed to see whether this is the case.

6 Conclusion

This study contributes to the knowledge on interaction with user-adaptive systems, and establishes the importance of transparency in interaction with user-adaptive systems. In our experiment, transparency increased the acceptance of recommendations. Even though in the current study users' acceptance of and trust in a content-based recommender system as a whole were not impacted by transparency, transparency can be considered an important aspect of interaction with user-adaptive systems. Findings show that the transparent version was perceived as more understandable and perceived understanding correlated with perceived competence, trust and acceptance of the system. Future research is necessary to evaluate the effects of transparency on trust in and acceptance of user-adaptive systems. A distinction has to be made in future research between acceptance of the system in users' context and acceptance of its results.

It is important that this work is replicated with different types of transparency and in various domains in order to come to a full understanding of the effects of transparency on user interaction and system acceptance. There is also a need to replicate the work in this paper in high-risk contexts in which the need for trust in a user-adaptive system may be more imperative. Trust in user-adaptive systems with (semi-) autonomous

functionality is scarcely evaluated. Even while for example an existing trust scale was adapted for use in this study, further research is needed in order to increase knowledge about trust and to develop reliable measurements that build on existing scales of trust.

Despite the limitations as identified in the discussion section above, the study has shown that transparency impacts acceptance of system results and that transparency features impact user behaviour in both expected and unexpected ways. This experiment provided empirical evidence on the influence of system transparency on acceptance of recommendations in a cultural heritage context. Our manipulation was successful and transparency increased user understanding and acceptance of the system's recommendations. The findings of this study argue for careful consideration of transparency features and explanations in user-adaptive systems. Beyond the positive effect of transparency on user interaction, this study offers evidence that different transparency features have different effects on user experience that need to be considered in the design of transparency features. Specific guidelines have been offered in this study for developing user-adaptive systems for cultural heritage contexts where specific application goals might require specific designs of personalisation strategies and explanations. Through this work, we also aimed to highlight some of the methodological challenges inherent in evaluating complex dimensions such as trust and contribute to the development of reliable scales for measuring acceptance of user-adaptive systems.

Acknowledgements We would like to thank all participants in this study and colleagues and anonymous reviewers for their helpful comments. 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). The Rijksmuseum Amsterdam gave permission for use of its artwork images.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alpert, S.R., Karat, J., Karat, C., Brodie, C., Vergo, J.G.: User attitudes regarding a user-adaptive e-Commerce web site. *User Model. User-Adapt. Interact.* **13**(4), 373–396 (2003)
- Amazon, www.amazon.com
- Aroyo, L., Wang, Y., Brussee, R., Gorgels, P., Rutledge, L., Stash, N.: Personalised Museum Experience: The Rijksmuseum Use Case. *The International Museums and the Web Conference, San Francisco, USA* (2007)
- Bartneck, C.: How convincing is Mr. Data's smile: affective expressions of machines. *User Model. User-Adapt. Interact.* **11**(4), 279–295 (2001)
- Benyon, D.: Adaptive systems: a solution to usability problems. *User Model. User-Adapt. Interact.* **3**(1), 65–87 (1993)
- Bilgic, M., Mooney, R.J.: Explaining Recommendations: Satisfaction vs. Promotion. *Beyond Personalization Workshop, The International Conference on Intelligent User Interfaces*, pp. 470–474. San Diego, California, USA (2005)
- Bowen, J.P., Filippini-Fantoni, S.: Personalization and the Web from a Museum Perspective. *The International Museums and the Web Conference*, pp. 63–78 (2004)
- Briggs, P., Simpson, B., De Angeli, A.: Trust and personalisation: a reciprocal relationship? In: Karat, C.-M., Blom, J., Karat, J. (eds.) *Designing Personalised User Experiences for e-Commerce*, pp. 39–55. Kluwer (2004)

- Brusilovsky, P.: Methods and techniques of adaptive hypermedia. *User Model. User-Adapt. Interact.* **6**(2–3), 87–129 (1996)
- Burke, R.: Hybrid recommender systems: survey and experiments. *User Model. User-Adapt. Interact.* **12**(4), 331–370 (2002)
- Carmagnola, R., Cena, F., Comsole, L., Cortassa, O., Gena, C., Goy, A., Torre, I.: Tag-based User Models for Social Multi-Device Adaptive Guides. This issue (2008)
- Carmichael, D., Kay, J., Kummerfeld, B., Niu, W.: Why did you show/tell/hide that? The need for scrutability in ubiquitous personalisation. ECHISE Workshop Exploiting Context Histories in Smart Environments at UbiComp, Irvine, CA, USA (2006)
- Castelfranchi, C., Falcone, R.: Trust and control: a dialectic link. *Appl. Artif. Intell. J.* **14**(8), 799–823 (2000)
- Cheverst, K., Byun, H.E., Fitton, D., Sas, C., Kray, C., Villar, N.: Exploring issues of user model transparency and proactive behaviour in an office environment control system. *User Model. User-Adapt. Interact.* **15**(3–4), 235–273 (2005)
- Cortellessa, G., Giuliani, M.V., Scopelliti, M., Cesta, A.: Key issues in interactive problem solving: an empirical investigation on users attitude. *Interact.*, pp. 657–670. Rome, Italy (2005)
- Cramer, H.S.M., Evers, V., Van Someren, M., Wielinga, B., Besselink, S., Rutledge, L., Stash, N., Aroyo, L.: User Interaction with User-Adaptive Information Filters, pp. 324–333. HCI International, Beijing, China (2007)
- Cramer, H.S.M., Wielinga, B.J., Evers, V., Rutledge, L., Stash, N.: The Effects of Transparency on Perceived and Actual Competence of a Content-Based Recommender. Semantic Web User Interaction workshop at CHI, Florence, Italy (2008)
- Damiano, R., Gena, C., Lombardo, V., Nunnari, F., Pizzo, A.: A stroll with carletto. Adaptation in Drama-Based Tours with Virtual Characters. This issue (2008)
- Davis, F.D.: Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quart.* **13**(2), 318–340 (1989)
- Dzindolet, M., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P.: The role of trust in automation reliance. *Int. J. Human-Comput. Stud.* **58**(6), 697–718 (2003)
- Fogg, B.J.: Prominence-interpretation theory: Explaining how people assess credibility online. Conference on Human Factors in Computing Systems CHI, pp. 722–723 (2003).
- Fogg, B.J., Tseng, H.: The Elements of Computer Credibility. Conference on Human Factors in Computing Systems CHI, pp. 80–87 (1999)
- Gefen, D., Karahanna, E., Straub, D.W.: Trust and TAM in online shopping: an integrated model. *MIS Quart.* **27**(1), 51–90 (2003)
- Goren-Bar, D., Graziola, I., Pianesi, F., Zancanaro, M.: The influence of personality factors on visitor attitudes towards adaptivity dimensions for mobile museum guides. *User Model. User-Adapt. Interact.* **16**(1), 31–62 (2006)
- Gregor, S., Benbasat, I.: Explanations from intelligent systems: theoretical foundations and implications for practice. *MIS Quart.* **23**(4), 497–530 (1999)
- Hanani, U., Shapira, B., Shoval, P.: Information filtering: overview of issues, research and systems. *User Model. User-Adapt. Interact.* **11**(3), 203–259 (2001)
- Herlocker, J.L., Konstan, J.A., Riedl, J.T.: Explaining Collaborative Filtering Recommendations. Conference on Computer Supported Cooperative Work CSCW, pp. 241–250. Philadelphia, Pennsylvania, USA, (2000)
- Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. *ACM Trans. Inform. Syst.* **22**(1), 5–53 (2004)
- Höök, K.: Evaluating the Utility and Usability of an Adaptive Hypermedia System. International Conference on Intelligent User Interfaces IUI, pp. 179–186. ACM, Orlando, Florida, USA (1997)
- Höök, K.: Steps to take before intelligent interfaces become real. *Interact. Comput.* **12**(4), 409–426 (2000)
- Höök, K., Karlgren, J., Waern, A., Dahlbck, N., Jansson, C.G., Karlgren, K., Lemaire, B.: A glass box approach to adaptive hypermedia. *User Model. User-Adapt. Interact.* **6**(2–3), 157–184 (1996)
- Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S.: MuseumFinland—Finnish Museums on the semantic web. *J. Web Semantics* **3**(2), 224–241 (2005)
- Jameson, A.: Adaptive interfaces and agents. In: Jacko, J., Sears, A. (eds.) *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pp. 305–330. Erlbaum, Mahwah, NJ (2003)
- Jameson, A., Schwarzkopf, E.: Pros and Cons of Controllability: An Empirical Study. *Adaptive Hypermedia and Adaptive Web-based Systems*, pp. 193–202. Springer (2002)

- Jian, J.Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn Ergonom.* **4**(1), 53–71 (2000)
- Jiang, X., Khasawneh, M.T., Master, R., Bowling, S.R., Gramopadhye, A.K., Melloy, B.J., Grimes, L.: Measurement of human trust in a hybrid inspection system based on signal detection theory measures. *Int. J. Ind. Ergonom.* **34**(5), 407–419 (2004)
- Jøsang, A., Lo Presti, S.: Analysing the relationship between risk and trust. *International Conference on Trust Management*, Oxford, UK (2004)
- Kurasaki, K.S.: Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field Methods* **12**(3), 179–194 (2000)
- Kay, J.: Scrutable adaptation: Because we can and must. *Adaptive Hypermedia and Adaptive Web-Based Systems AH*, pp. 11–19. Dublin, Ireland (2006)
- Kim, S., Alani, H., Hall, W., Lewis, P., Millard, D.E., Shadbolt, N.G., Weal, M.J.: Artequakt: Generating Tailored Biographies with Automatically Annotated Fragments from the Web. *Semantic Authoring, Annotation and Knowledge Markup Workshop at ECAI*, pp. 1–6. Riva del Garda, Italy (2006)
- Klopping, I.M., McKinney, E.: Extending the technology acceptance model and the task-technology fit model to consumer e-commerce. *Inform. Technol. Learn. Perform. J.* **22**(1), 35–48 (2004)
- Langheinrich, M.: Privacy by Design—Principles for Privacy-Aware Ubiquitous Systems. *International Symposium on Ubiquitous Computing Ubicomp*, pp. 273–291. Atlanta, GA (2001)
- Lee, J.D., Moray, N.: Trust, self-confidence, and operator's adaptation to automation. *Int. J. Hum.-Comput. Stud.* **40**, 153–184 (1994)
- Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. *Hum. Factors* **42**(1), 50–80 (2004)
- Ma, Q., Liu, L.: The technology acceptance model: a meta-analysis of empirical findings. *J. Organ. End User Comput.* **16**(1), 59–72 (2004)
- McAllister, D.J.: Affect and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad. Manage. J.* **38**(1), 24–59 (1995)
- McGuinness, D.L., Pinheiro da Silva, P.: Explaining answers from the semantic web: the inference web approach. *Web Semantics. Sci. Serv. Agent World Wide Web* **1**(4), 397–413 (2004)
- McNee, S.M., Lam, S.K., Guetzlaff, C., Konstan, J.A., Riedl, J.: Confidence Metrics and Displays in Recommender Systems. *Interact.*, pp. 176–183. Zurich, Switzerland (2003)
- McSherry, D.: Explanation in recommender systems. *Artif. Intell. Rev.* **24**, 179–197 (2005)
- Muir, B.M.: Trust in automation: part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* **37**, 1905–1922 (1994)
- 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)
- Ndubisi, N.O., Gupta, O.K., Ndubisi, G.C.: The Moguls' model of computing: integrating the moderating impact of users' persona into the technology acceptance model. *J. Glob. Inform. Technol. Manage.* **8**(1), 27–47 (2005)
- Nückles, M., Winter, A., Wittwer, J., Herbert, M., Hübner, S.: How do experts adapt their explanations to a layperson's knowledge in asynchronous communication? An experimental study. *User Model. User-Adapt. Interact.* **16**(2), 87–127 (2006)
- Oppermann, R., Specht, M.: A nomadic information system for adaptive exhibition guidance. *Arch. Museum Inform.* **13**(2), 127–138 (1999)
- Pandora, www.pandora.com
- Parasuraman, R., Miller, C.: Trust and etiquette in high-criticality automated systems. *Commun. ACM* **47**(4), 51–55 (2004)
- Pavlou, P.A.: Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. *Int. J. Electron. Comm.* **7**(3), 101–134 (2003)
- Proctor, N., Tellis, C.: The state of the art in museum handhelds in 2003. *Museums and the Web*. Charlotte, NC, USA (2003)
- Picard R.W.: *Affective Computing*. MIT Press, Cambridge, MA, USA (1997)
- Pu, P., Chen, L.: Trust-inspiring explanation interfaces for recommender systems. *Knowl.-Based Syst. J.* **20**, 542–556 (2007)
- Rajaonah, B., Anceaux, F., Vienne, F.: Trust and the use of adaptive cruise control: a study of a cut-in situation. *Cogn. Technol. Work* **8**(2), 146–155 (2006)
- Rajaonah, B., Anceaux, F., Vienne, F.: Study of driver trust during cooperation with adaptive cruise control. *Le Travail Humain* **69**, 99–127 (2006)

- Rajaonah, B., Tricot, N., Anceaux, F., Millot, P.: Role of intervening variables in driver-ACC cooperation. *Int. J. Hum. Comput. Stud.* **66**(3), 185–197 (2008)
- Reeves, B., Nass, C.: *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press, Cambridge, UK (1996)
- Rocchi, C., Stock, O., Zancanaro, M., Kruppa, M., Krueger, A.: The Museum Visit: Generating Seamless Personalized Presentations on Multiple Devices. *International Conference on Intelligent User Interfaces*, pp. 316–318. Madeira, Portugal (2004)
- Schreiber, G., Amin, A., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Hollink, L., Huang, Z., van Kersen, J., de Niet, M., Omelayenko, B., van Ossenbruggen, J., Siebes, R., Taekema, J., Wielemaker, J., Wielinga, B.J.: *Multimedial e-culture demonstrator.* *International Semantic Web Conference*, pp. 951–958. Athens, USA (2006)
- Shimazu, H.: ExpertClerk: a conversational case-based reasoning tool for developing salesclerk agents in e-commerce webshops. *Artif. Intell. Rev.* **18**, 223–244 (2002)
- Shneiderman, B., Maes, P.: Direct manipulation vs. interface agents. *Interactions* **4**(6), 42–61 (2007)
- Simmel, G.: In: Wolff, K.H. (ed.) *The Sociology of George Simmel.* Free Press, New York (1964)
- Sinclair, P., Lewis, P., Martinez, K., Addis, M., Prideaux, D.: *Semantic Web Integration of Cultural Heritage Sources.* 15th International Conference on World Wide Web, pp. 1047–1048. New York, USA (2006)
- Sinha, R., Swearingen, K.: The role of transparency in recommender systems. *Conference on Human Factors in Computing Systems*, pp. 830–831. ACM Press (2002)
- Stock, O., Zancanaro, M., Busetta, P., Callaway, C., Krüger, A., Kruppa, M., Kuflik, T., Not, E., Rocchi, C.: Adaptive, intelligent presentation of information for the museum visitor in PEACH. *User-Model. User-Adapt. Interact.* **17**(3), 257–304 (2007)
- Tintarev, N., Masthoff, J.: Effective Explanations of Recommendations: User-Centered Design. *ACM Conference on Recommender systems*, pp. 153–156 (2007)
- Trant, J.: Exploring the potential for social tagging and folksonomy in art museums: proof of concept. *New Rev. Hypermedia Multimedia* **12**(1), 83–105 (2006)
- van der Heijden, H.: User acceptance of hedonic information systems. *MIS Quart.* **28**, 695–704 (2004)
- Van Setten, M.: *Supporting People in finding information: hybrid recommender systems and goal-based structuring.* Telematica Instituut, The Netherlands (2005)
- Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. *MIS Quart.* **27**(3), 425–478 (2003)
- Victor, P., Cornelis, C., De Cock, M., Pinheiro da Silva, P.: Gradual trust and distrust in recommender systems. To appear in *Fuzzy Set. Syst.* (2008)
- Waern, A.: User involvement in automatic filtering: an experimental study. *User Model. User-Adapt. Interact.* **14**(2–3), 201–237 (2004)
- Wang, Y.: *The Presentation of Media-rich Collections of Culture Heritage in the Age of Digital Reproduction.* Vrije Universiteit, Amsterdam, The Netherlands (2005)
- Wang, W., Benbasat, I.: Trust in and adoption of online recommendation agents. *J. Assoc. Inform. Syst.* **6**(3), 72–101 (2005)
- Wang, W., Benbasat, I.: Recommendation agents for electronic commerce: effects of explanation facilities on trusting beliefs. *J. Manage. Inform. Syst.* **23**(4), 217–246 (2007)
- Wang, Y., Aroyo, L., Stash, N., Rutledge, L.: Interactive User Modeling for Personalised Access to Museum Collections: The Rijksmuseum Case Study. *International User Modeling Conference UM*, pp. 385–389. Corfu, Greece (2007)
- Wu, I.L., Chen, J.L.: An extension of trust and TAM model with TPB in the initial adoption of on-line tax: an empirical study. *Int. J. Hum.-Comput. Stud.* **62**(6), 784–808 (2005)
- Wubs, H., Huysmans, F.: *Snuffelen en Graven.* Sociaal en Cultureel Planbureau, The Netherlands (2006)
- Xiao, S., Benbasat, I.: The formation of trust and distrust in recommendation agents in repeated interactions: a process-tracing analysis. *The 5th international conference on Electronic commerce*, pp. 287–293 (2003)
- Zimmermann, A., Lorenz, A.: LISTEN: a User-Adaptive Audio-Augmented Museum Guide. This issue (2008)

Author's vitae

Henriette Cramer is a PhD-candidate at the University of Amsterdam's Human-Computer Studies lab. She received her Master's degree in Social Science Informatics from the same university in 2004, specialising in human-computer interaction. Henriette's current Ph.D. research focuses on user acceptance and trust of user-adaptive and (semi-)autonomous systems. Her additional interests include human-robot interaction, context-awareness, affective computing and mixed-reality systems. Her previous research includes work on interaction with Virtual Reality systems.

Vanessa Evers holds a MSc from the University of Amsterdam and a PhD from the Open University, UK on Cultural aspects of Human Computer Interaction. Vanessa conducts research about cultural issues of interaction and interaction with intelligent systems. She has been involved in the organization of the Workshop on International Issues for Products and Systems since 1999, has been the editor of the Designing for Global Markets proceedings and is a member of the organising committee of the Human Robot Interaction international conference.

Satyan Ramlal received his master degree in Social Science Informatics from the University of Amsterdam, under the supervision of Dr. Vanessa Evers. His interests lie in the role that individual and organisational contexts play in the ways in which information technology is shaped and accepted by end-users. Satyan contributed to this study as part of his Master's graduation project.

Maarten van Someren is a Lecturer in Artificial Intelligence at the University of Amsterdam. He has worked in several areas of artificial intelligence, natural language processing and cognitive modelling including machine learning, knowledge-based systems, user modelling, models of human learning and cognitive development. He has authored over a hundred technical papers and has edited several books.

Lloyd Rutledge is an Assistant Professor at the Open University Netherlands. He received his B.S. in Computer Science from the University of Massachusetts Amherst and his M.S. and Sc.D. in Computer Science from the University of Massachusetts Lowell. He has also worked as a researcher at CWI, the Dutch National Center for Mathematics and Computer Science. His research topics there ranged from authoring Web-based multimedia to automatic generation and personalisation of presentations using Semantic Web technologies.

Natalia Stash received her PhD from the Eindhoven University of Technology (TU/e), The Netherlands. She is currently a scientific programmer in the CHIP project (Cultural Heritage Information Personalisation) between the Rijksmuseum (Amsterdam), TU/e and Telematica Institute (Enschede). Her research interests include adaptive hypermedia, semantic web technologies and learning styles.

Lora Aroyo is Assistant Professor of Computer Science at the VU University Amsterdam, working in the area of semantic web and user modeling. She completed her PhD in 2001 at the University of Twente, in the area of intelligent educational systems. Her recent research concentrates on ontological support for knowledge intensive work, including cultural heritage, interactive TV and e-learning. She is scientific coordinator of the CHIP Project in collaboration with Rijksmuseum Amsterdam. The joint research described in this volume reflects an interest in applying personalisation techniques in semantic enriched museum collection for enhancing users' experience.

Bob Wielinga studied physics at the University of Amsterdam, where he was awarded a PhD degree cum laude in 1972 for a thesis in nuclear physics. In 1977, Wielinga was appointed senior lecturer at the Department of Psychology of the University of Amsterdam. Since 1983, Wielinga has performed research on the methodology of knowledge-based system design and knowledge acquisition. In 1986, Wielinga was appointed full professor of Social Science Informatics in the Faculty of Psychology. In this capacity Wielinga was and is team leader of several research projects, including KADS, ACKnowledge, REFLECT and KADS-II. He was one of the main contributors to the development of the KADS methodology for knowledge based system development.