People's responses to autonomous and adaptive systems
Cramer, H.S.M.

Citation for published version (APA):

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: http://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.
8 CONCLUSIONS AND DISCUSSION

This thesis has studied interaction between users and adaptive and autonomous systems in various contexts. The main goal of this thesis was to investigate the influence of understanding-related, as well as social-affective aspects of interaction on user perceptions of, trust in and behaviour toward adaptive and autonomous systems. An additional goal was to identify interaction breakdowns and challenges, and to see how various interaction design features can impact the interaction between users and systems that to a certain extent make their own decisions.

The previous chapters have described studies on interaction with spam filters in an everyday setting, effects of transparency in interaction with a user-adaptive recommender system, a video-based survey study on in-vehicle agents, two video-based survey studies on social behaviour of physically embodied robots and a Wizard-of-Oz experiment investigating the effects of socially expressive, empathic behaviour on mobile interaction with a system that monitors the user’s surroundings.

This chapter revisits the research questions posed in the introduction. It summarises the research endeavours described in this thesis and their main results. Implications for the design of adaptive and autonomous systems that interact with users are described, as well as the limitations of the approach taken in this thesis. Directions for future research and a discussion of the challenges ahead conclude this chapter.

8.1 SUMMARY OF RESULTS

The system characteristics that have been investigated in these studies clearly affect user perceptions and trust. However, the specific effects of these characteristics are not clear cut and turn out to be dependent on the interplay of user, system and context characteristics. The consequence of this general observation is that future research and design of interaction features of autonomous systems cannot be performed in a vacuum; we have to consider what the effects of these features are in different circumstances and for different people. Having said that, it becomes clear from the various studies in this thesis that adaptive, autonomous systems differ from traditional ICT applications in that users are much less sure about the functionalities and capabilities, feedback requirements and intentions of these systems. This uncertainty results in suboptimal use and user behaviour that can be unexpected or even erratic at times. As the adaptive systems adapt themselves to the user, users in turn also adapt their behaviour to their perceptions of the
system’s inner workings. When confronted with results users think are unsuitable or which they do not understand, they will start experimenting with the system to both understand it and to gain better results. While experimenting is not necessarily a negative thing, we do need to consider that users will for example tailor their feedback in a manner that they think will yield them the best results, but in actuality only decreases performance.

Although attitudes toward adaptive, autonomous systems are often positive, users do not always sufficiently feel they can depend on systems enough to delegate tasks to them, and this applies to both low and high risk situations. Trust is a multi-faceted concept and simple attempts to increase trust by social and empathic interaction styles or by adding transparency features will only in very specific settings yield their expected results.

Another important finding in a number of our studies is the role of congruency between system behaviour and the user’s expectations and context. Incongruence negatively affects trust in most contexts. Negative effects for example occur when social behaviour is inappropriate considering the situation at hand. Incongruence of the users’ expectations of the effect that their feedback will have on a system and the actual results a system produces leads to less trust and unexpected user behaviour. Social behaviour can be incongruent with the level of autonomy of a system and finally, incongruence between personality traits of the user and the ‘personality’ traits and behaviour of a system can also have a negative effect. Below the results of the studies are discussed following the research questions posed in the introduction.

8.1.1 How do awareness and understanding affect perceptions, attitudes, trust, delegation, feedback and training toward a system?

Awareness and understanding of adaptive and autonomous systems is not a trivial matter. The study on interaction with spam filters (Chapter 2) shows that even when interacting with a familiar and relatively simple adaptive system that potentially can make autonomous decisions, user awareness of the system’s activities and inner workings can vary widely. In the recommender study (Chapter 3) participants’ perceptions of the non-transparent system’s inner workings did not necessarily match the actual criteria the recommendations were generated with, nor did user perceptions of the system’s competence necessarily match its actual competence. The impact of incomplete understanding was especially striking in situations where users need to provide information to a system. Misconceptions about the training process led to sub-optimal feedback behaviour, as illustrated by incomplete training of the spam filter in Chapter 2. In the mobile study (Chapter 7), user responses to the system’s requests depended on their perceptions of what type of information the system needed and what type of input it could deal with. Even highly educated participants in the mobile study were insecure about what capabilities adaptive and autonomous systems could have. Participants questioned how the system worked and what the reasons were for its messages. For some participants this was a reason
for not fully complying to the system’s requests and warnings. For others it sparked curiosity to find out how the system worked.

The relationship between (perceived) understanding and perceptions of a system’s competence, dependability and acceptability is not clear-cut. User understanding does not necessarily lead to more user trust. In the recommender study in Chapter 3, perceived understanding was related to perceived competence, trust and acceptance of the system. However, trust in, and acceptance of, the system itself was not improved by making the system more transparent. Explaining to the user why a recommendation was made, however, did increase acceptance of the individual recommendations. In the spam study (Chapter 2), understanding was not related to whether participants thought they could rely on their filter. Transparency and increased understanding about a system’s inner workings can increase acceptance, but only if users perceive the transparent system results as fitting for their purposes and the offered information on the system’s inner workings appears useful to users for such judgment calls. A full understanding is not a guarantee for more system acceptance; a system’s results and inner workings need to be up to the user’s standards. If a system’s criteria for its decisions are transparent, it is also easier for users to detect incongruencies with the way they think a decision should be made.

These results show that system designers need to pay special attention to ensuring awareness about systems’ activity and adaptivity. A full understanding of a system’s inner workings might not be feasible, or even necessary. Chapter 3 however showed that making a system more transparent can at least increase the chances that user perceptions of competence and actual competence of a system match. System developers must make sure that users at least have a level of understanding that allows them to make an accurate assessment of a system’s competence and that ensures users know what type of feedback a system needs to adapt and function correctly.

8.1.2 How do the level of system autonomy and user control affect perceptions, attitudes, trust, compliance, feedback and training toward a system?

The studies in this thesis yielded insight on both the effects of the level of autonomy of a system, as well as the effects of user control in the form of training and feedback. Both are discussed below.

8.1.2.1 User feedback and training

The spam filter study (Chapter 2) shows that adaptivity and user investment in training do not necessarily lead to more trust in a system. Trust in the training process instead was found to play an important role in the user’s willingness to invest effort in a system that can be taught to improve over time. Users need to be able to assess what the results are of their feedback. They also need guidance on what type of information to provide. If users notice system mistakes, it is important they can directly correct that specific mistake. In the recommender study’s transparency condition (Chapter 3) for example the level of transparency did not match the type
of feedback users could provide, this led to suboptimal feedback from users trying to correct mistakes in an indirect manner. The possibility for backtracking by users and undoing the results of their previous feedback need to be available as users can be apprehensive of training when they are not fully sure of what the result of their feedback will be. At the same time - in tension with the need for such direct user control - safeguards have to be built into systems that rely on user feedback to ensure suboptimal feedback or inaccurate information provided by users, do not decrease performance.

8.1.2.2 System autonomy

Even when users have a general trust that a system will perform its task well, they are not necessarily willing to assign a high level of autonomy to a system in their specific context (see e.g. Chapters 2, 3, 4). The level of control users have over the system has to be congruent with the situation. Before people actually delegate tasks to a system a number of conditions have to be fulfilled.

First of all, a trade-off between risks and the perceived utility of a system is made; the level of perceived risk has to be acceptable (see e.g. Chapters 4 and 2). In high risk situations the advantages of delegating to a system are overridden by the perceived risks.

When we change the level of autonomy to fit a certain context, we also have to consider the expectations raised by autonomous behaviour. System autonomy can for example affect the type of (social) behaviour that is deemed suitable, or even expected from a system (such as in Chapter 5).

We need to consider that raising the level of control does not raise trust in the collaboration with a system in every context. Individual differences and personality characteristics also affect the level of autonomy that is accepted from a system.

8.1.3 What is the effect of systems’ social expressive behaviours such as empathy on perceptions, attitudes, trust and compliance?

Chapters 5, 6 and 7 explicitly addressed the effects of systems’ social behaviours on user interaction with them.

Social behaviours can affect the perceptions of users and their attitudes towards systems. Social behaviour can have important positive effects; in the study on the effects of empathic accuracy on interaction with social robots (Chapter 6) for example we found that dependability, credibility and closeness increase when a robot shows accurate empathic behaviour. In Chapter 5 we found that touch for proactive robots increased the perception of the robot being less like a machine.

However, (inappropriate) social behaviour also can have clear negative consequences. The results of our studies emphasise the importance of congruency of displayed behaviours with the interaction context.

The results from especially Chapter 6 and 7 illustrate the importance to evaluate the effects of social, empathic behaviour and not just to investigate whether social behaviour is recognised. Both studies also show the influence of how social
abilities are conceptualised and their subsequent design and evaluation. Simply making system messages seem friendly, (blunt) statements on the user’s experience and shallow mimicking of social behaviours does not constitute a more ‘natural’ interaction or empathy with the user’s experience. In Chapter 6 we found that accurate social behaviour and reacting in line with the user’s affective experience is important. Dependability, credibility and closeness decrease when a system shows behaviour incongruent with the situation. However, in negative situations, users can attribute an agent greater empathic abilities when an agent, incongruent with the negative situation, remains optimistic and this behaviour is perceived as an attempt to make the user feel better. In Chapter 7 a more expressive mobile system was seen as more empathic than a neutral version, but this did not increase trust in the system and did not increase compliance with its requests and information it provided. Rather the behaviour of the system appeared considered as socially awkward by many participants and incongruent with the urgency of the situation. Coupled with the interruption of the user’s task, this had a consequence that most users considered the system statements as insincere and inappropriate. The way a message is delivered needs to be congruent to its content and the usage context.

Social behaviours cannot be evaluated in isolation; their effects depend on the context the behaviours are shown in. Social behaviours that can be evaluated positively in one domain, can be considered inappropriate or distracting in other domains. Social behaviour also have to fit a systems’ overall behaviour. In the study on the effects of touch in interaction with social robots (Chapters 5) we found that touch by a proactive robot has a positive effect, while for a reactive agent the absence of touch resulted in a greater perception of dependability. Both of these combinations can be viewed as a manifestation of congruency; proactive systems are more expected to interact using social behaviours than reactive systems.

Even though we observed in Chapter 6 that in negative situations showing positive empathic behaviour was considered a sign of empathic abilities, the mobile study in Chapter 7 shows that the manner in which a message is delivered still has to be congruent with its content. The increased perceptions of sociability when a system is overly optimistic in a negative situation do however bring up ethical concerns related to misuse of social system behaviours. Blind optimism from a system might for example distract users from negative situations and mistakes. It would be interesting to see whether more subtle behaviours that are more realistic about the current situation could yield positive results.

Whether intentionality is attributed to a system and whether perceived goals behind system’s requests coincide with the user’s current goals are additional factors that could affect whether expressiveness is experienced as a ‘sincere’ positive feature of user-system dialogues. A clear distinction has to be made between the effects of social behaviours on trust in a system overall and effects for specific advice and requests. Perceived empathy in Chapter 7 for example was correlated with perceived dependability of a system and its source credibility, but not directly to trust in the provided information. Careful consideration of which behaviour will fit the context, user and system (both its purposes and its form) is crucial.
8.1.4 What are the effects of differences between individual users and differences in interaction context (urgency, emotional valence of a situation) on perceptions, attitudes, trust and compliance?

8.1.4.1 Effects of individual differences

Individual differences, personality characteristics, personal opinions and knowledge can affect whether a system is appreciated. We have seen that personality traits such as locus of control (Chapter 4), empathic tendencies (Chapter 7) and attitudes toward a particular type of technology (Chapters 5 and 6) affect responses to a system. People with an internal locus of control are more likely to accept a higher level of autonomy from a system. General attitudes towards specific types of technology affect whether social behaviours are accepted from a system.

Reactions to social system behaviours also depend, at least partially, on whether a system’s interaction style matches the user’s personality. Whether for example socially expressive behaviours are appreciated and how they are perceived depends on personal preferences and opinions. Users’ personalities affect the interaction style and the effect that the level of sociability has on their affective experience (e.g. Chapter 7). Users with a lower level of empathy for example, will not respond more positively when a system is more empathic. Delegation to a system also depends on individuals’ assessment of their own ability to perform a task, coupled with their level of expectations on how a system would perform. Doubts about techniques used in a system can result in negative attitudes towards a system, especially for users with expert knowledge (specifically in Chapters 2 and 3). Individual differences can also have an indirect effect on whether using a system is acceptable, for example through the differences in perceptions of the context a system is used in (Chapter 4). Design and evaluation should take individual differences that may occur in the target user population very seriously indeed.

8.1.4.2 Contextual differences

The spam filter study showed that trust is not necessarily the determinant of delegation of tasks to a system. The context in which a system is used instead leads to a situational assessment whether a user can depend on a system. Factors such as situational risks, urgency (Chapters 4, 7) and consequences of system mistakes (Chapter 2) appear to play a dominating role in user attitudes, delegation and compliance.

We have to consider the situatedness of users’ behaviour and responses to a system. How users respond to system requests or whether they comply to its advice will depend on what they experience in their specific surroundings (e.g. Chapter 7). The results of Chapter 6 show that valence of the situation in which users interact with a system, also affect perceived credibility and dependability of an agent. This poses quite a challenge for developers of systems used in situations in which task success and a pleasant affective experience are not a given such as in healthcare, or crisis management.
The situational context also plays a role in the perception of social behaviours. In Chapter 6 emotional valence of a situation interacted with the effects of empathic accuracy; in more negative situations the robot was seen as more empathic when it (inaccurately) remained positive about the situation. However, in Chapter 7 socially expressive behaviour for instance appeared to be at odds with the serious nature of hazard monitoring and warnings.

8.2 LIMITATIONS

This thesis has provided a broad overview of issues that arise when people interact with autonomous and adaptive systems in various domains. This provides useful information for system designers on the range of interaction issues that they need to consider and for researchers on research issues that deserve future attention. However, this also means the issues have been highlighted, but not solved. Currently little guidance is available for system designers on the subtleties involved and how to successfully implement interaction aspects for adaptive and autonomous systems. The guidelines proposed in various places in this thesis can provide direction and inspiration for future research.

The methodological approach taken in this study of individual user studies in various domains also has its limitations. All studies in this thesis involved specific settings with individual users. There might be limitations to the generalisability of these results to other settings. Also, not all studies in this thesis explored usage of systems outside of the lab. In the majority of the studies, participants had not encountered the system in question before. This means they had to familiarise themselves with the system or agent they were evaluating during their session and that they were building a first mental model of its workings. In settings where users interact with systems for a longer period and a relationship has been built, aspects such as social behaviour or transparency of system behaviour might have different effects. Some studies also only used video-scenarios. This has the advantage that more futuristic systems and simulated usage scenarios can be explored, but also the disadvantage of a more distant evaluation of a system without users actually experiencing an interaction themselves.

The experiments in this thesis have been developed from a point of view based on specific theories, such as the media equation theory, and on for example an approach to trust that incorporates both a cognitive and social, affective standpoint. Such perspectives naturally affect how a study is designed, which issues are salient during for example observation, the way results are explained and which directions for further research and system development are proposed. Exploring the issues highlighted in this thesis from different perspectives will provide an even richer overview.
8.3 IMPLICATIONS FOR FUTURE RESEARCH

From both the above studies’ results and limitations a number of implications can be drawn for future studies. There are methodological challenges inherent in evaluating complex concepts such as trust and acceptance of a system. The results of for example Chapter 2 show that the level of reported trust, does not necessarily match the level of delegation. Clear distinctions have to be made between effects of interaction features on for example trust in a system overall and effects for trust in specific advice or requests and compliance. Trust and acceptance should be investigated on a number of levels. We have shown that trust as a general attitude towards a system independent of its use in a specific situation has to be distinguished from situational trust and reliance on a system in a specific context. In turn, trust in the results that a system produces can be different from general and situational trust. The plausibility of specific system decisions, answers or recommendations will again affect whether they are followed-up upon. These nuances are often not captured in research on interaction with adaptive and autonomous systems. Trust-related scales, such as the dependability and credibility scales used in this thesis, can be adapted, but further research is needed in order to increase knowledge about trust and to develop reliable measurements that build on such existing scales and metrics. A lack of shared evaluation criteria and measurements for trust and a clear definition of what constitutes a user’s experience, makes comparisons over studies in the field difficult. Clear conceptualisations and validated scales can help assess the effects of interaction in a more comparable manner.

We should take care not to try and capture the user’s experience in rigid, numerical criteria only. Complementing quantitative measures with qualitative and behavioural exploration is especially useful as for example reported trust in our studies did not match user behaviour. The spam filter study showed that people indicated they trusted the system when presented with a numerical rating scale, while observation indicated they did not ‘implement’ this reported trust in their behaviour. A similar issue was found in the mobile study (Chapter 7) on social expressiveness; participants rated the system as empathic, but in the interviews some of them actually indicated they thought the system was sarcastic, insincere or unfit for the context. Rich analysis using ethnographic techniques and more open observation and interviews can yield insights in why users have certain attitudes and show certain behaviour that never can be captured in a score on a scale (Höök, 2004). This is especially important for autonomous and adaptive systems as complex mental models, social-affective reactions and unexpected behaviours can be invoked by system behaviour.

Studies investigating interaction with adaptive and autonomous systems should take into account participant personality when analysing the results of interaction features. Some aspects of a user-system dialogue might be beneficial for some users, while being detrimental to the experience of others. Ignoring such possible differences will lead to a shallower understanding of how interaction design features can impact attitudes and trust. Further analysis of personality effects would be very
useful, especially since current information on the effects of personality traits on interaction with autonomous systems is still fragmented.

Social behaviours are also heavily cultural dependent and research results might not universally apply. The relative importance of cognitive aspects of user interaction and social aspects of interaction will most likely also be context-dependent. We also have to consider that when users interact with systems for a longer time and a relationship has been built, reactions to social behaviours might be different.

It’s important to realise that adaptive and autonomous systems will come in very different forms. They range from ‘traditional desktop programs’, to embodied robot creatures, to distant services interacting via user’s own digital devices, to sensor-outfitted environments. How embodiment will for instance affect persuasiveness of system decisions and recommendations is unclear.

The field of HCI is constantly moving and overarching theories are still relatively scarce. The inherent interdisciplinary nature of HCI makes it important for HCI researchers and system developers to gain from the expertise available in other fields. However, the risk of cherry-picking aspect from other disciplines that might appear promising, without a full understanding of their underlying principles is relatively large. A deeper understanding of the circumstances in which these principles might apply is necessary.

8.4 CHALLENGES

The findings in this thesis point to a number of challenges for both developers and researchers. First of all, the difficulties of users to predict what an autonomous or adaptive system will do in a given situation can have serious consequences. The differences in perceptions of our participants and the actual behaviour and inner workings of the systems they were evaluating, highlight the risk that users over- or underestimate a system’s capabilities. This poses great difficulties in deciding when usage is or is not appropriate. We cannot always predict how technologies will be implemented, how they will be used and what their consequences will be and this especially applies to systems that learn, change and show autonomous behaviour. Science fiction has presented us with ample horror scenarios of autonomous systems gone haywire. While such visions might not paint the most realistic picture, in some cases they do serve as a useful reminder. We at least need to make sure that we try to gain insight in the consequences of these technological developments and realise that adaptive and autonomous systems are not just a distant vision. We lack a full insight in how people interact with autonomous and adaptive systems and which aspects influence this behaviour, even for those systems that are already widely used. This is a huge problem, especially since autonomous and adaptive systems are being developed for critical domains with potential life and death consequences. Unfortunately, discussions on ethics, safety and privacy are moving slower than technology developments itself.

Research on social implications of adaptive and autonomous agents is yet scarce. Adaptive and semi-autonomous agents will also affect social relations in a more
direct way. People recommenders for example arrange social contact (such as Facebook.com’s friend and activity suggestions). Foucault et al. (2007) for example showed that a ‘gossiping’ agent spreading rumours it had gathered from its users, caused a positive, increased sociability between its users. However, when autonomous systems will mediate between people more negative effects and misunderstandings are not unimaginable.

This also points to another challenge: responsibility for a system’s actions. From a more traditional point of view, system mistakes can be seen as user error, putting responsibility of the user, or as a design flaw putting the blame on the system developer. For autonomous and adaptive systems this is much less clear. Where responsibility lies when a system makes a mistake when users have trained a system is not a simple question, especially when they might have understood the system differently than designed. It is unclear who will be responsible for the actions of systems that learn, change and make decisions in ways that developers might not have foreseen. Systems may appear to develop their own goals and intentions; how these are related to end-users’ goals might be hard to identify. It may be hard to establish whether a system is acting on behalf of an individual (legal) person or organisation and whether they can and should be held responsible for a system’s actions.

An additional challenge is posed by the need to pay more attention to affective experiences of users, which will especially affect the user’s interaction when a full understanding of a system is not possible. When such a full understanding is not possible, social-affective aspects of the interaction will likely be the most likely factor to persuade (or dissuade) users to trust a system. The work in this thesis has shown that these aspects and their relation to trust are in now way straightforward, but a challenging avenue for further research.

Directions suggested in the literature cannot be directly applied in every situation. Before embarking on a mission to make a system more adaptive or more autonomous, it is worthwhile to consider whether the benefits of such adaptive and autonomous behaviour are likely to outweigh their potential downsides in the specific situation at hand, or whether other solutions would be more suitable. Such a consideration also applies to interaction features when actually developing an adaptive or autonomous system in situations where they have clear potential. Features such as transparency will not always lead to increased trust, acceptance and optimal use. The social behaviours suggested to ease interaction with autonomous systems cannot be blindly implemented in every system and their effects will depend on the context and user. Social behaviours and features related to user feedback and system explanations cannot be ‘tacked on’ to a system as an afterthought. They need to be an integral part of its interaction design and need to be designed and evaluated in a situated manner. It’s very easy to mess up, and much harder to get behaviours right.

Adaptive and autonomous systems offer great promises for the role of technology in the future, but their nature provides us with complex challenges. This work has attempted to identify these challenges and has laid a basis for addressing these issues in an informed way.