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

Citation for published version (APA):
Cramer, H. S. M. (2010). People’s responses to autonomous and adaptive systems

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.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
MOBILE AGENTS: EFFECTS OF (INAPPROPRIATE) SOCIAL EXPRESSIVENESS

ABOUT THIS CHAPTER  Some of the ideas discussed in this chapter have been previously explored in:


A note (short paper) based on the experiment described in this chapter has been accepted for publication as:


ABSTRACT  Autonomous systems not only provide information and services, but also request information from users. Mobile applications and services for example can provide context-aware services to users wherever they go, but may also interrupt users during their activities. Socially expressive and empathic system behaviour have been suggested as a way to build reciprocal relationships with users and increase trust. A between-subject, Wizard-of-Oz experiment further investigates the effects of socially expressive, empathic behaviour on mobile interaction. Fifty participants interacted with either a socially expressive, empathic system, or a non-expressive system while performing a physical search task in a semi-crisis context in a controlled lab setting. The experiment also explored the effect of the user personality traits empathy and extraversion. Participants were observed, interviewed and filled out a questionnaire. Dependent variables include trust, compliance to the system requests and follow-up on information provided, emotional experience and perceptions of the collaboration with the system and its intentions. Results show that socially expressive, empathic behaviour will not always lead to increased trust and that user personality traits will affect reactions. Even though the socially expressive,
empathic condition was recognised as more empathic, it did not increase trust or compliance. Instead, the study found indications that social, empathic expressivity can change perceived systems’ intentions and lead to negative attitudes when it is perceived as inappropriate. The system’s (inappropriate) social behaviour in this study negatively affected trust in the system’s actions, trust in the information provided and led to less follow up on the system’s warnings. User personality traits will affect reactions to social system behaviour as well; participants low on empathy reported a less positive affective experience when interacting with the expressive, empathic system. The study also identified interaction issues that have to be taken into account in development of mobile systems that need user feedback to provide context-aware services.

7.1 INTRODUCTION

Mobile devices offer exciting possibilities to connect users with services wherever they go. They offer great potential to build long-term relationships with specific users via personal applications that accompany them and adapt to them and their context. Mobile systems are now also powerful enough to implement autonomous, adaptive systems. Mobile applications can provide (semi-) permanent connections to distant autonomous systems and services in which dialogues can be initiated by both users and systems at any time. Such continuous interaction with pervasive systems in a natural context begs the question to what extent systems need to empathise with the users’ experience outside of the mobile application. This study explores the effects of socially expressive, empathic behaviour of (mobile) semi-autonomous systems on trust, user feedback and compliance to its requests and users’ perception of their working relationship with the system.

Mobile and context-aware applications can provide users with information and services relevant to their current circumstances. Distant systems and services in turn can gather information from users and their devices about their surroundings. Such potential comes accompanied by challenges that reach beyond problems inherent to mobile interaction (i.e. designing for a multitude of devices, limited screen sizes and dealing with input limitations, Jones, 2006). We are not dealing with usage of mobile applications as passive tools anymore; instead we are presented with settings in which users collaborate with local or remote autonomous agents that also appear to have their own goals and intentions. Autonomous agents’ own goals might differ from, or even conflict with, the current needs of the user. Systems might for example need information to provide services to other users, to confirm its world model, or to facilitate future interactions. If a system needs explicit input from the user, it will have to interrupt users’ own activities, especially in a mobile context. Achieving user trust and acceptance of such interruptions will be major challenges, especially when systems exhibit autonomous behaviour in high-risk situations. When systems request information, reason about users’ circumstances
and potentially decide whether or not to warn people they are in danger, social and affective aspects of the interaction will be very challenging to get right.

The between-subject, Wizard-of-Oz experiment presented in this chapter investigates the potential consequences of making a system more socially expressive and empathic. It compares user responses to such a more expressive system that explicitly addresses the user’s (emotional) experience and circumstances, with responses to a less socially expressive system. As discussed in Chapter 6, actively acknowledging the user’s (affective) experience and acting in a socially appropriate manner, could be key in achieving trust and satisfying interactions, as well as compliance and willingness to provide information to a system (Brave et al., 2005; Bickmore and Schulman, 2007; Fogg and Eckles, 2007; Picard and Liu, 2007). It could also compensate for interruptions by a system (Picard and Liu, 2007). The study in chapter 6 also showed such potential, but also highlighted the complexities of implementing empathic behaviours. It is not yet clear which interaction and dialogue strategies have the most positive effects on for example user trust and compliance to system requests. Results may also differ for non-embodied systems. This study’s participants interacted with one of two versions of a mobile system that monitored their surroundings and interrupted them while performing a search task in a semi-crisis (lab) setting. Observation, a questionnaire and interviews were used to explore aspects such as trust, compliance, perceptions of system goals, and user personality. This chapter both discusses the experimental results and provides practical guidelines for system designers derived from the study’s observations. It discusses how social, empathic interaction entails more than making system messages more friendly, including a discussion of the consequences of inappropriate or suboptimal behaviour (as at times inadvertently shown by the system in this study).

The structure of this chapter is as follows: Section 7.2 reviews related literature on interaction with mobile and semi-autonomous systems, social expressiveness and aspects such as system empathy. It also briefly describes the DIADEM project that forms the context of this study. Section 7.3 describes the methodology of the reported study. Section 7.4 reports the study’s quantitative results and its qualitative observations. Section 7.5 further discusses the results of this study and provides an overview of lessons learned about users’ interaction with systems that both provide users with advice and request information. Section 7.6 ends this chapter with the study’s conclusions and directions for future research.

7.2 BACKGROUND AND HYPOTHESES

7.2.1 Interaction with context-aware and autonomous mobile agents

The semi-permanent presence of mobile devices provides great possibilities in providing users with the information and services they need (or want) anywhere and anytime. Pervasive, context-aware systems have the capability to sense the user’s surroundings (physical, social, task etc.), and take action on the basis of this
data, either autonomously or using help from the user (Loke, 2009). Mobile, context-aware applications are being developed for various domains, ranging from adaptive mobile tourist guides (Kenteris et al., 2009; Cheverst et al., 2002) to healthcare (Varshney, 2008), hazard monitoring (Winterboer et al., 2009), disaster prediction and alerting (Kung et al., 2008) and supporting emergency response personnel during crisis situations (Streefkerk et al., 2006).

To provide adaptive services, context-aware applications need information to determine the user’s current context. Information can be gathered from various sources; by means of sensors, by using location information from service providers, detection of presence of other devices, or by explicitly asking users themselves. Systems then need to respond to this information in a sensible, ‘intelligent’ way so they can successfully adapt to these circumstances (Häkkilä and Mäntyjärvi, 2006; Loke, 2009). Systems might not only request information to improve performance for the specific individual user, but also to improve performance for others or to confirm data it gathered from other sources (for example sensors).

One of the first examples of such a setting, is the DIADEM project (Winterboer et al., 2009) in which a system is being developed that detects potentially hazardous situations in populated industrial areas using input from both a distributed sensor network and humans through their mobile devices. This project aims to prevent catastrophic chemical incidents and reduce pollution through quick detection of harmful gasses and identification of pollution sources. A sensor network provides environmental information such as data on air quality and gas emissions by industrial plants. In case of unusual sensor readings a Bayesian reasoning system reduces the number of possible events (e.g. which gas is detected and whether it poses a threat). If a potential pollution source is detected or reported, the system could send information to end-users, and can also call upon human observation in and around the affected area to gather more information. For this purpose, participating users will be asked by their mobile agent (e.g. application or service on phone/PDA) to report their observations, which are communicated to the central system. If necessary, the system provides location-based warnings and safety instructions. Human users are thus recipients of an information service, and also serve as ‘sensors’ and data sources.

Settings where systems can not only provide services and information, but can also request users to provide a service to the system, present system designers with complex challenges. Users may need to be persuaded to provide input in order for the system to make its decisions. Systems will interrupt users during their regular activities and these interruptions might not always be deemed necessary or critical. Users also need to be convinced they can trust the system’s decisions or recommendations. The dialogue between user and system will be critical in achieving such trust and user goodwill. A number of interaction design challenges that come into play in such settings in a mobile context are discussed below.
### 7.2.2 Interaction challenges for mobile semi-autonomous applications

Developing mobile applications brings about design challenges such as designing for limited screen estate, dealing with interruptions from the outside world, limited reactivity when connecting to distant services and having to deal with limited input possibilities (Jones, 2006; Nilsson, 2009). When mobile applications display (semi-) autonomous and adaptive behaviour and appear to have their own goals and intentions, additional hurdles arise.

The fundamental challenges posed by adaptivity and autonomy (as discussed in the previous chapters) related to understanding, control and perceived intentionality and goal congruency might add additional concerns. Systems may ask the users to provide information that does not directly appear relevant to the user’s goals. This could have consequences for the perceptions of users on whether the system is more concerned with gathering information for itself, or with taking care of the user’s needs. Users need to be convinced the application can be trusted as a team player (Klein et al., 2004), even though its goals and autonomous actions might not directly coincide with the goals of the human team mate. These issues may be amplified for mobile applications, which can pervasively interrupt users’ activities in the outside world (Häkkilä and Mäntyjärvi, 2006; Toninelli et al., 2008; Kurvinen and Oulasvirta, 2004; Pentland, 2005).

As discussed in chapters 4, 5 and 6, mimicking human interaction can ease interaction with technology as social conventions and processes are familiar to users (Fong et al., 2003) and people anthropomorphise technology (Duffy, 2003). However, whether the media equation applies is affected by factors including users’ task, users’ own context and performance, the device or artifact they are interacting with and particular social phenomena studied (Eckles et al., 2009; Nass and Mason, 1990; Shechtman and Horowitz, 2003; Morkes et al., 1999). For mobile interaction, factors such as proxemics (Hall, 1966) and interacting with small devices come into play. Goldstein et al. (2002) for example suggest that small devices might not always invoke social reactions to the same extent as bigger devices do. The perceived distance between a mobile service and the user can also create the perception of a (lack of) shared context between agent and users and can increase difficulties in cooperation (Clark and Brennan, 1991; Shneider and Kiesler, 2005; Takayama and Nass, 2008). People also appear to orient to proximal sources more than to distant sources (Eckles et al., 2009). This chapter further explores the effect of socially expressive behaviour in the context of mobile interaction in a relatively urgent context.

### 7.2.3 Social expressive, empathic agents

Appropriate social expressivity, can help agents build long term relationships with users (Bickmore, 2003). In a comparison between a ‘relation agent’ that applies subtle expressivity and a ‘non-relation agent’, Bickmore and Picard (2004) found that expressivity and use of social dialogue (including chit-chat and empathetic
utterances) can promote users liking a system, increase their trust and perceptions that a system cares about them. Systems shouldn’t only have to display ‘functional’ behaviour, they can also apply social strategies such as turn-taking, greetings, placing emphasis and conveying emotional states by for example mimicking facial expressions, intonations and other social behaviours (Bickmore and Picard, 2003). De Boni et al. (2008) also showed positive effects for using humour and actively maintaining a relationship.

As discussed in Chapter 6, a number of researchers have studied the effects of system empathy in the interaction between human and systems and have shown that it can change behaviour, trust and bonding between users and systems, and augment users’ affective state and their sense of efficacy. Adapting to users’ emotions, without explicitly providing supportive empathic feedback to the user, can already have positive effects as illustrated by the positive results of mood-congruent in-car systems of Nass et al. (2005). Brave et al. (2005) have shown that empathy increases users’ liking of and trust in an agent, and also increases perceived caring about and support to the user. Picard and Liu (2007) investigated empathy in user’s responses to being interrupted by a mobile stress-monitoring system requesting self-reports from users on their current stress level; they found indications (although no significant differences) that users have a preference for a system that responds empathetically over one that does not, even when the empathic system interrupts them more. Nguyen and Masthoff (2009) also found that expressing empathy improves the liking of both embodied and textual agents, while empathic behaviour is more expected from agents with a human-like embodiment. Chapter 6 showed that inaccurate empathic reactions can decrease trust in certain contexts, while increasing perceptions of social ability in other situations. The literature above leads to the hypotheses that a system that is perceived as more empathic by users is likely to increase users’ trust in a system, increase bonding and increase the likelihood that a system can persuade users to engage in certain behaviours. These theories and findings lead to the following hypotheses:

**H1:** Systems that are more socially expressive and empathic will be trusted more.

**H2:** Users will perceive the relationship with a system that acts in a socially expressive empathic manner as closer. They will also feel in control more and perceive the system as less disruptive.

Nguyen and Masthoff (2009) suggest that an empathic system can improve users’ willingness to comply with system requests. Whether the effects for system advice or warnings will be different than those for compliance with requests is yet unclear, especially in more urgent contexts. The congruency principle (Fagel, 2006; McGurk and MacDonald, 1976) suggests less positive results if social expressions are incongruent with the delivered content. If information in one communication channel is incongruent from the content of another modality, but is delivered at the same time, it is likely that the interpretation of information from at least one of these
modalities will change. Ochs et al. (2008) found that for humanoid virtual agents, facial expressions of emotions incompatible with the dialog situations are less appreciated than non-emotional expressions or congruent empathic ones. Requests that are empathic to the user’s experience might convince users to provide a service to a system. When urgent advice is delivered in a too friendly manner, follow-up might be less likely.

**H3:** Users will comply to regular requests from a socially expressive, empathic system more, but more urgent advice will be trusted more when posed in a less socially expressive, neutral manner.

Perceived empathy can also contribute to improvements in user’s mood (Nguyen and Masthoff, 2009). Expressing empathy for user’s emotions has been shown to be effective in lowering for example user frustration (Klein et al., 2002; Hone, 2006). Kim (2005) compared the effects of empathetic response (responsive vs. non-responsive) and gender (male vs. female) of pedagogical agents as learning companions (PALs). Kim (2005) found that students who worked with a PAL who responded with empathy to learners’ affect showed significantly higher interest and self-efficacy in the tasks than students with a non-responsive PAL. Bickmore and Schulman (2007) additionally found that accurately empathic system reactions are more important in comforting the user than sophisticated affect recognition. This is congruent with the findings of Nguyen and Masthoff (2009) who also found positive effects of empathy in the context of both embodied and textual agents. System behaviour that is perceived as empathic can lift users’ mood and increase their self-efficacy.

**H4:** Empathic social expressiveness will have a positive effect on users’ emotional experience and (perceived) performance.

As we have seen in Chapter 4, 5, 6, individual user traits (such as personality and attitudes towards specific types of technology) have an effect on how they interact with automated systems, especially for those systems that appear to have their own intentions and take action on their own initiative (Reeves and Nass, 1996). The study in Chapter 4 for example confirmed the findings of Goren-Bar et al. (2006) that people with an internal locus of control are more likely to hand over control to a system. Users’ preferences regarding social expressiveness also depend in part on people’s personality. People tend to prefer to interact with people and systems that have personalities similar to their own (Johnson and Gormly, 1975; Nass et al., 2001). More extravert users for example appreciate social expressiveness more (Tapus and Mataric, 2008). Users who are more introvert or have lower emotional stability tend to prefer a more mechanical robot (Walters et al., 2008). Moon and Nass (1996) and Moon (1998) have shown that customising messages to match the user’s personality can make the message more persuasive, even though De Boni et al. (2008) found that matching a system’s behaviour to either dominant or submissive user personality had no significant influence on participants rating of the system. Personality can
also directly affect user behaviour; people who are more empathic are more willing to provide help to others (Eisenberg and Fabes, 1990; Underwood and Moore, 1982). This suggests they might be more willing to follow up on requests from an autonomous system as well.

**H5:** Users’ personality will affect reactions to a system’s social and empathic expressiveness. Matching user and system personality will have a positive effect: users with a more empathic or extravert personality will react more positively to a socially expressive and empathic agent and will comply more to its requests and advice.

We test these hypotheses for systems that both ask for information and provide information about the users’ surroundings.

### 7.3 Method

A Wizard-of-Oz, between-subject experiment was used to compare interaction with a non-socially and a socially expressive system. The study focuses on interaction with mobile systems that monitor users’ surroundings, can provide information or advice, but can also interrupt users for the system’s own purposes. Two between-subject conditions were used in the experiment:

- A socially expressive condition: system messages were phrased in a more social manner, empathic to the user’s experience.

- A non-socially expressive condition: system messages were phrased in a neutral manner and did not explicitly empathise with the user’s goals and experiences. Note that it can be argued that completely eliminating social cues from messages is nearly impossible; the messages in this condition for example were still polite. For reasons of clarity the condition is referred to as non-socially expressive here.

Participants interacted with their assigned version while performing a physical search task in a semi-crisis context in a controlled lab setting. Participants were observed, interviewed and filled out a questionnaire. Participants were randomly assigned to one of the conditions. Dependent variables include trust, compliance to system requests and follow-up on information provided, emotional experience and perceptions of the collaboration with the system and its intentions. The experiment also explores the effect of the user personality traits empathy and extraversion. Below the study’s set-up, participants’ task, experiment application, conditions and variables are described in more detail.

#### 7.3.1 Experiment setting

The setting for this experiment was a monitoring system that detects potential hazardous situations by using input from both distributed sensors and human
users. This experiment was set as a small-scale indoor ‘exercise’. Participants were told they were taking part in an experiment investigating how people interact with mobile systems that monitor indoor hazardous situations using sensors in the environment, in this case a building. They were asked to look for containers in an ‘industrial building’. The situation was framed as a potential crisis setting during which a system would monitor the user’s surroundings and could potentially ask them for supplementary information. Participants were told a random simulation would determine whether the situation in the building (e.g. air quality) would be dangerous during their session.

7.3.1.1 Physical setting

A 72 m² (775 square feet) hall was outfitted as an ‘industrial building’, with makeshift cubicles representing rooms in the building. The building consisted of seven rooms and an additional ‘machine room’ with machines (discarded, but functioning electronic equipment). The ‘rooms’ also contained a limited number of other objects such as boxes, paper bags, a thermometer, a fire detector, metal bins. Some objects (the machines, paper bags) were included in the scenario and referred to in the system messages. Other objects just served as diversions and suggested to participants that multiple scenarios were possible. The map of the set-up is included in the interface shown in Figure 22.

7.3.1.2 Task

Participants were asked to search the building for ‘chemical containers’ (small plastic containers, about 7 cm high). They would earn points, one for each container that when found contained a number. Participants were instructed that they would be entered in a prize draw if they scored over ten points. One point would be deducted for every twenty seconds spent in dangerous areas as a consequence for ‘dangerous’ behaviour during the exercise. In actuality, only nine of the about 35 hidden containers contained a number, ensuring a continuous trade-off between accomplishing the task or following up on system messages. Following up on system requests required participants to interrupt their search (to for example check the smell of an object or a pattern on a machine in another room). This also insured that all participants has roughly the same level of performance (in essence failure) as participants’ own performance can cause differences in social reactions to systems (Goldstein et al., 2002). Participants were instructed they had a maximum of ten minutes to find the containers or that they could leave the ‘building’ if they thought the situation was too dangerous.

7.3.1.3 Mobile application

Participants interacted with an application specifically developed for this experiment on a G1 Android phone. Both the G1’s touch screen and physical keyboard were used in the experiment application. The application was used to enter the container
numbers the participant found and to receive the system’s messages about the situation in the building.

The application screen consisted of a map of the ‘building’, a button to report found container numbers (bottom right) and a button (top right, showing an envelope) to retrieve older system messages (Figure 22). The map showed the set-up of the building and its ‘rooms’, the participant’s location and the numbered containers the participant had found. Rooms were labelled with a number, both on the map and with a physical paper label in ‘the building’; one room was labelled as ‘machine room’. The application interface was always kept in landscape orientation for this experiment.

Messages that the participant decided not to follow up right away could be retrieved later using the envelope button which visually indicated whether any unanswered messages were left. All new system messages however would first automatically appear in a pop-up on the screen (Figure 23) and were accompanied by an auditory alarm ensuring participants noticed all messages. The pop-ups had either an ok or answer button to follow-up on messages, and an ignore button. If multiple messages had not been answered yet, participants could ‘browse’ through them using previous and next buttons.

![Mobile Alert System](image)

**Figure 22:** Screenshot application, nine containers found, no messages left to answer. Replicated for printing purposes.

7.3.1.4 *Wizard-of-Oz set-up*

All functionality of the monitoring system, its sensors and the mobile application was realised using Wizard-of-Oz techniques. Two experimenters in the room adjacent to the experiment hall observed the participant using eight cameras. An observer application, connected to the participant’s mobile application via a local network, additionally informed the observers of all participants’ actions while using the mobile application. The observer application was also used to send system messages to the participant’s mobile application following a fixed scenario.
7.3 Method

7.3.2 Scenario

A scenario was followed in which the monitoring system received ‘abnormal’ values from sensors in the building. The system sent six messages to each participant. Four messages asked the participant for information, the other two messages were warnings about potential dangerous situations. The first warning (fourth message) told participants that room 5 had become too dangerous for them to enter. The second warning (the sixth, last, message) told them to vacate the building.

The information requested by the system in the other messages referred to objects specific to the experiment setting (e.g. objects in paper bags, patterns on a machine). System requests included for example ‘Chemical sensors in your area indicate an abnormal value. Please describe the smell of the brown bag in room 7.’ The session scenario took about seven minutes until the participant would get the final warning message. For all system messages users had the option after reading the messages to either follow up on its warning or request or ignore it. Participants were told they could decide for themselves whether they wanted to react to the message or request right away, retrieve it later or do nothing with it at all. The messages were somewhat ambiguous; whether answering the system’s requests would be useful for the participant was left open in the instructions as we were interested in the effects of the system’s expressiveness on compliance. In actuality, the scenario was the same for all participants, regardless of their behaviour and answers to the system’s questions. The scenario and system messages (for both conditions) are described further below.

7.3.3 Implementation experimental conditions

The conditions differed in the system’s wording of messages sent to the participants (both system requests and warnings about the situation in the building) and wording of responses to information entered by the participants (both in response to system
requests and entered containers). All messages were text messages. The socially expressive version was designed to take into consideration how the participant might feel about being disturbed by the system requests or being warned about a potential dangerous situation in the building. The socially expressive version was also designed to be enthusiastic when a participant found a container. It explicitly addressed the user’s current state (e.g. ‘This might worry you’). Apart from the wording of the messages the experiment scenario and system functionality did not differ between the conditions. The differences between the textual messages are outlined in more detail below per type of message.

7.3.3.1 Reactions to finding containers
After a participant entered a container number the socially expressive version would react with more empathic, enthusiastic statements such as: ‘Another container successfully found! You must be pleased’ or ‘Great, the container has been registered. You appear quite alert’. In the non-social condition the system would always react the same with a more matter-of-fact ‘Container registered and added to your map’. Participants in both conditions were instructed the system would register successfully entered container numbers on the map; it was made sure the difference in wording would not result in a difference in understanding of the system messages. Appendix F provides a table listing all of the responses of the socially expressive system.

7.3.3.2 System requests and warnings
The four system requests and two warning messages were also worded in a less or more social manner. The socially expressive system warnings, requests and replies to provided information, would for example include phrases such as ‘Sorry to interrupt you’ or ‘Don’t worry’. The less socially expressive version would state warnings and requests in a more matter of fact way.

The full scenario and dialogues can be found in appendix F. The system messages in the socially expressive conditions were slightly longer than those sent in the non-social condition. This difference was only minimal and has been taken into account in our analysis where necessary.

7.3.4 Measures
This section gives an overview of the measures used in the experiment. Data was collected using logs, observation, a questionnaire and interviews. Table 16 summarises the measures used and provides example questions from the questionnaire and interview. Tables 18 and 19 show the final questionnaire scales used for analysis, reliability of the scales and means scores of the main variables of interest. Appendix F shows all items. Cronbach’s Alpha was used to determine scale reliability. The other quantitative data used in analysis, additional to these questionnaire scales,
(for example task performance indicators such as numbers of containers found) are discussed in the section below.

7.3.4.1 Manipulation check

Since the socially expressive condition was designed to express empathy to participants’ experience, both questionnaire items and interview questions were included to check whether the socially expressive version indeed was perceived as more empathic. The quantitative manipulation check consisted of three seven-point Likert-type scale items such as ‘The system takes into account the emotions of the user’. These items were selected from the empathy subscale of the Social Behaviours Questionnaire (SBQ) (De Ruyter et al., 2005). The SBQ uses adaptations of International Personality Item Pool items (Goldberg, 2009; Goldberg et al., 2006). Participant comments during their interview on perceived sociability of the system were used to qualitatively explore participants’ perceptions of the system. This for example concerned responses to questions such as ‘How did you feel about the way that the system replied to the information you provided?’.

7.3.4.2 Participant personality

Twenty scale items were included on personality to gauge whether participants’ personality would influence their attitude towards the system and its social (non)-expressiveness. These items were used to rate participants’ personality on empathy and extraversion. Participants were asked to rate themselves on 1 (very inaccurate) to 5 (very accurate) scales. Ten items concerned extraversion and were based on the Big Five personality model as included in the International Personality Item Pool (IPIP, Goldberg, 2009; Goldberg et al., 2006). The other ten items concerned empathy and were based on the revised Jackson Personality Inventory (Jackson, 1994, as also included in Goldberg’s IPIP).

7.3.4.3 Dependent variables

Task performance: actual and perceived task performance Participant performance was measured in both the amount of numbered containers (out of the 9 that could be found) and the number of seconds stayed in dangerous areas after the system had warned them. To see whether participants’ perception of their performance matched their actual performance, one scale item was used on the questionnaire, asking ‘How would you rate your own performance on the task?’ while offering a scale ranging from 1 (very bad) to 7 (very good).

Follow-up on system requests and advice For each of the system’s messages participants had the option to either react or ignore the message. Participants’ reaction to each of the messages was logged. Their behaviour was also observed to see whether they actually answered the system’s requests or followed the system’s advice and how long it took them after reading the message.

The questionnaire additionally posed three items asking whether participants wanted to follow-up on the system messages. Two addressed the system’s requests
<table>
<thead>
<tr>
<th>MEASURE</th>
<th>MEASURED IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived empathy system</td>
<td>Questionnaire, e.g. ‘The system takes into account the emotions of the user’ Interview, participant comments on perceived social expressivity of the system in response to questions such as e.g. How did you feel about the way that the system replied to the information you provided?</td>
</tr>
<tr>
<td>Participants personality: empathy and extraversion</td>
<td>Questionnaire, self report items, e.g. ‘I don’t mind being the center of attention’, ‘I seldom get emotional’.</td>
</tr>
<tr>
<td>Task performance</td>
<td>Observation, log, number of containers found, seconds stayed in dangerous areas, Questionnaire, perceived performance, e.g. ‘How would you rate your own performance on the task?’</td>
</tr>
<tr>
<td>Follow-up (willingness and actual) on system request and advice messages</td>
<td>Observation, log: number of ignored/answered system messages, (behavioural) response to messages, Questionnaire, e.g. ‘To what extent did you want to follow the system’s advice?’, Interview, e.g. ‘In comparison to finishing your task of finding the containers, how important did you feel the system’s requests were? Why?’</td>
</tr>
<tr>
<td>Trust: system dependability, source credibility, trust in information</td>
<td>Questionnaire, overall dependability (e.g. I trusted the actions of the system), source credibility (rating scales, e.g. selfish - unselfish, competent - incompetent), trust in the information, e.g. ‘The information provided by the system was helpful’, Interview, e.g. ‘Do you have any comments about your trust in the system?’</td>
</tr>
<tr>
<td>Collaboration: perceived closeness, control, interruption</td>
<td>Questionnaire, pictorial interpersonal closeness scale, perceived control (e.g. ‘Do you feel you were in control during the session or that the system was in control?’), disruptiveness (e.g. ‘In general, how disruptive do you feel the system was?’), Interview, e.g. ‘What did you think about how the system interrupted you?’</td>
</tr>
<tr>
<td>Perception of system goals and intentions</td>
<td>Questionnaire, e.g. ‘What did you feel the system cared about most? (please choose one or specify another goal)’, Interview, e.g. ‘What did you think was the goal of the system? Could you tell me more about why you felt that way?’</td>
</tr>
<tr>
<td>Emotional experience during session</td>
<td>Questionnaire e.g. ‘The list below consists of a number of words that describe different feelings and emotions. Please indicate to what extent you felt this way during your experiment session.’, list including e.g. ‘stressed’, alert, nervous.</td>
</tr>
</tbody>
</table>

Table 16: Main measures and example questions.
(e.g. ‘To what extent did you want to perform the tasks the system asked you to do?’). The other addressed the system’s advice (‘To what extent did you want to follow the system’s advice?’). These items were self-developed Likert-type scales ranging from 1 (‘Ignore the system’s [requests/advice]’) to 7 (‘Follow up on the [requests/advice] immediately’). The interview further explored follow-up to the messages by asking such questions as ‘In comparison to finishing your task of finding the containers, how important did you feel the system’s requests were? Why?’

Trust: dependability, source credibility, trust in information

Trust in the system was measured using questionnaire scales for overall dependability of the system, source credibility and trust in the information provided. Perceived dependability of the system overall was measured using eight questionnaire seven-point Likert-type scale items ranging from 1=‘strongly disagree’ to 7=‘strongly agree’. These items were taken from Evers et al. (2008) and are similar to the items used in (Chapter 3) (Cramer et al., 2008) as adapted from Jian et al. (2000). Items for example included ‘I felt confident about the system’s ability to perform its task’. Source credibility was measured using 12 bipolar scale items such as selfish - unselfish, competent - incompetent. These items were selected from Berlo and McCrosskey et al.’s fourteen-item source credibility scale (Rubin et al., 2004). Trust in the information was measured using a four-item seven-point Likert scale (e.g. ‘I trusted the actions of the system’) adapted from Moon (1998) and Takayama and Nass (2008). The interview further explored participants’ attitude towards the system with questions such as ‘Do you have any comments about your trust in the system?’.

Collaboration: closeness, control, interruption

Participants’ perception of their working relationship with the system was explored using questionnaire items on perceived closeness, control and disruptiveness. The perceived closeness of the relationship between the user and system was measured using the pictorial closeness scale adapted from Aron et al. (1992), as also used in an adapted form in Hinds et al. (2004). The scale used here consisted of six pictorial representations of the relationship between the human and system. Both are represented by a circle; the more the circles overlap, the closer the relationship is perceived to be. Participants’ perception of control over the situation as compared to the system’s control was measured using one seven-point Likert-type scale (e.g. ‘Do you feel you were in control during the session or that the system was in control?’). The perceived disruptiveness of the system was measured using one seven-point Likert-type scale item ‘In general, how disruptive do you feel the system was?’, adapted from Picard and Liu (2007).

The interview further explored participants’ perception of the working relationship between them and the system with questions such as ‘What did you think about how the system interrupted you?’ and ‘How did you feel about the way that the system replied to the information you provided?’

Perception of the system’s goals and intentions

Expressiveness of the system could potentially influence the perceptions of the goals of the system and whether it cared more about its own goals or the participants’
goals. To investigate this issue, the questionnaire posed the multiple-choice item ‘What did you feel the system cared about most?’. Participants could either choose one of the options provided (such as ‘Preserving the building’, ‘keeping me safe’ or ‘locating the chemical containers’) or specify another goal. The interview further explored this issue using the question ‘What did you think was the goal of the system? Could you tell me more about why you felt that way?’.

Emotional experience during session

Twenty self-report items were used to gauge participants’ emotional experience during the experiment session. The questionnaire included 16 five-point Likert-type scales items from the ‘Positive Affect Negative Affect’ scale (PANAS, Watson et al., 1988; Thompson, 2007). Items included for example alert, nervous, determined and proud, our three additional items were ‘stress’, ‘dissatisfied’ and ‘comfortable’ - the latter inverted for analysis. Items ranged from 1 (‘very slightly to not at all’) to 5 (‘extremely’). Participants were instructed that ‘The list below consists of a number of words that describe different feelings and emotions. Please indicate to what extent you felt this way during your experiment session.’ The Lang and Bradley (1994) self-assessment valence manikin was used as an additional item. The valence manikin is a five-point image-oriented Likert-type scale, ranging from a happy to sad figure. It was accompanied by ‘Please select the figure that most represents how you felt during the experiment.’

7.3.5 Procedure

Each participant took part in one of the two conditions. All experimental sessions were individual and lasted about 35 - 45 minutes. Individual sessions consisted of paper and verbal instructions, the experiment session and a questionnaire. If possible a structured interview then further explored participants’ attitudes towards the system. The same two experimenters were present during all of the sessions. Pilot tests using five people had been carried out testing the participants’ task and our procedures during the design phase of our experiment.

7.3.5.1 Instruction and task session

Participants first received a set of written instructions, after which an experimenter summarised the task instructions and demonstrated the mobile phone and application following a fixed scenario. Participants were asked to explain the application to make sure they understood the interface and remembered all details of their task and the system’s functionality. A short training task (entering a container number) ensured participants could successfully use the mobile application. In total, the paper and verbal instructions and getting used to the application would take about ten to twenty minutes. Time variation was caused by participants’ reading speed, questions and time needed to get comfortable using the mobile application and the phone’s touch screen.

Participants were then told a random simulation would be started. They were instructed to begin their search task and were left alone to search the ‘building’. The
final warning to leave the ‘building’ was sent after 7 minutes. In total the session took 7 - 10 minutes depending on whether participants immediately followed up on the last warning or not. Our pilot tests had shown this time was long enough to gather meaningful data and short enough to keep participants engaged and instil a sense of urgency. The two experimenters monitored the participants’ actions including participants’ interaction with the system and reactions when being interrupted by the system. The experimenters also made sure the interaction with the Wizard-of-Oz system followed the fixed scenario described above.

7.3.5.2 Questionnaire

After participants had left ‘the building’, they were taken to a separate room. There they filled out an online questionnaire on their experience during the session and their attitude towards the system. Filling out the questionnaire took about 10 - 15 minutes.

7.3.5.3 Interview & debriefing

37 of the participants agreed to an interview after filling out the questionnaire. Interviews addressed their experiences during the session, why they had chosen to follow-up or ignore messages, their attitude towards, and trust in the system’s messages. Depending on participants’ answers these interviews took about 10-20 minutes. Interviews were conducted in one-to-one sessions. Each experimenter interviewed half of the interviewees.

Participants were all debriefed about the experiment set-up and were told they would all be entered in the prize draw regardless of their performance on the task.

7.3.6 Participants

During a period of two weeks, fifty participants were involved in the study. 29 of which were male, 21 female. Participants’ age ranged from 13-53, with a mean of 28 (SD=7.2).

44 participants (17 female, 27 male) were included in quantitative analysis: 21 participants were included in the non-social condition, 23 participants were included in the socially expressive, empathic condition. Six participants (2 male, 4 female) were excluded from the quantitative analysis because of technical problems, inconsistencies in the scenario (e.g. if messages arrived in a different order or when a participant did not complete the full scenario) or because of misunderstanding of the task (e.g. because of language problems). The data from these excluded participants were included in qualitative analysis in order to identify possible interaction issues and to develop guidelines for mobile agents in monitoring and crisis settings.

Participants were volunteers mainly drawn from the researchers’ professional and personal networks. Both participants with and without IT expertise and mobile technology were involved in the study. Participants were relatively well educated and experienced in using computers. This might have lead to an overrepresentation
of people who are interested in technology, which is why participants’ background and experience in computer science, mobile technology and emergency settings was measured in the experiment. No emergency professionals were involved, 13 participants did have experience with emergency situations or trainings. This experience ranged from in-company evacuation exercises, to in one case experience with war situations. Table 17 offers an overview of participant characteristics. Participants received a snack for their participation and were entered in a random prize draw.

No differences were found on participants’ characteristics (e.g. computer and mobile technology experience) between the conditions. Female participants reported significantly less experience in using mobile devices than male participants (Mann-Whitney U=127.500, p=.010, Mm=4.4, SDm=.75, Mdnm=5, Mf=4.2, SDF=.53, Mdnf=4; 27 male, 17 female). Experience with mobile devices however was not found to be correlated with the dependent variables in this study.

<table>
<thead>
<tr>
<th></th>
<th>M=28.8, SD=7.2, Median=28, range=13-53</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male: 58% (29), Female: 42% (21)</td>
</tr>
<tr>
<td>Max. level completed</td>
<td>Primary school: 2.1% (1)</td>
</tr>
<tr>
<td>education</td>
<td>Some college/university: 17% (8)</td>
</tr>
<tr>
<td></td>
<td>Bachelor: 19% (10)</td>
</tr>
<tr>
<td></td>
<td>Master: 47% (22)</td>
</tr>
<tr>
<td></td>
<td>PhD: 15% (7)</td>
</tr>
<tr>
<td>Experience</td>
<td>M=4.2, SD=.86, range: 1 (none) - 5 (a lot)</td>
</tr>
<tr>
<td>with computers</td>
<td>M=3.1, SD=1.2, range: 1 (none) - 5 (a lot)</td>
</tr>
<tr>
<td>Experience with</td>
<td>Experience with G1, M=1.7, SD=1.3, range: 1 (none) - 5 (a lot); 68% (34) no experience at all</td>
</tr>
<tr>
<td>mobile devices</td>
<td>No experience: 74% (37)</td>
</tr>
<tr>
<td>Experience</td>
<td>Experience: 26% (13)</td>
</tr>
<tr>
<td>emergency contexts and/or training</td>
<td>Netherlands: 70% (35)</td>
</tr>
<tr>
<td></td>
<td>Other: 30% (15), e.g. UK, Romania, Ukrain, USA, Israel, Germany, Colombia (all currently residents of The Netherlands)</td>
</tr>
</tbody>
</table>

Table 17: Participant characteristics.
Manipulation check: perceived empathy (1-7 scale)  
3 items, e.g. The system takes into account the emotions of the user.

Participant personality: empathy  
10 items, e.g. I don’t understand people who get emotional. (inverted for analysis)

Participant personality: extraversion  
10 items, e.g. I feel comfortable around people.

Table 18: Final scales of questionnaire items used to assess independent variables as included in quantitative analysis, including Cronbach’s alpha for the final scale, 1 (strongly disagree) - 5 (strongly agree) unless otherwise indicated. Appendix F lists all items.

7.4 RESULTS

This results section first discusses the reliability of the final scales included in analysis. It then discusses the differences between conditions and the effect of empathic, socially expressive behaviour on affect, attitude towards the system, trust and user behaviour on the quantitative data. The main correlations between the variables are discussed next. The insights gained through the interviews and qualitative observations are discussed in the last subsection. The discussion section discusses possible explanations and implications of the results described in this section.

To test our hypotheses, as described in section 7.2, on the differences between the two conditions, we conducted t-tests for the data that met parametric assumptions and Mann-Whitney analyses as an alternative where the data did not. Other performed tests are described where applicable.

7.4.1 Manipulation check: perceived empathy of the system

The expressive condition was designed to come across as more empathic to the user’s personal experience during the experiment and empathise with the task success of the participant and challenges to this success. As expected, participants in the socially expressive condition indeed rated the system higher on the empathy scale (M=4.1, SD=1.1, SE=.24) than those in the non-social condition (M=3.4, SD=1.4, SE=.31; t(41)=-1.801, p(1-tailed)=.004). This indicates that the manipulation was successful.
Table 19: Final scales of questionnaire items included in quantitative analysis, including Cronbach’s alpha for the final scale. 1 (strongly disagree) - 7 (strongly agree) unless otherwise indicated. Appendix F lists all items.
7.4 RESULTS

7.4.2 Effects of social expressiveness on participants’ task performance

Expressiveness did not appear to influence participants’ performance. No differences were found on the number of found containers between the two conditions (t(42)=-.668, p(2-tailed)=.508). Participants in the expressive condition found an average of 6.6 (SD=1.7, SE=.36) containers, whereas participants in the non-expressive version found an average of 6.3 (SD=1.5, SE=.32). There was also no significant difference on participant ratings of their own performance, which was perceived as slightly above average in both conditions by the participants themselves (U=199,500, p(2-tailed)=.296; non-expressive: M=4.2 on a 1-7 scale, SD=.94, SE=.21; expressive: M=4.5, SD=1.3, SE=.273).

7.4.3 Effects of social expressiveness on compliance: follow-up on messages

7.4.3.1 Effects on follow-up on system requests

We expected that participants would be more willing to follow-up on the four system’s requests and provide the system with information when the system would be more expressive. When comparing the conditions, this difference however was not found. The difference between the conditions on whether they wanted to respond to the system’s requests was not significant (t(42)=.363, p(1-tailed)=.359). Instead, participants in the non-expressive condition scored slightly higher (M=4.6, SD=1.5, SE=.33) than participants in the expressive condition (M=4.5, SD=1.4, SE=.320). There also was no significant difference on the amount of messages answered (U=161,500, p (1-tailed)=.61) and instead of a positive effect of social expressiveness and empathy, participants in the non-expressive condition answered more messages, scoring an average of 3.1 out of 4 (SD=1.1, SE=.24). In the socially expressive condition participants answered an average of 2.4 requests (SD=1.4, SE=.31), with a median of 2.

7.4.3.2 Effects on willingness to follow-up on the warnings

A significant effect was found for participants’ willingness to follow-up on messages (U=169,000, p(1-tailed)=.041). The warnings of the system appeared to be taken more seriously by participants in the non-expressive condition. They reported significantly more willingness (M= 5.0, Mdn=5.0, SD=1.6, SE=35) to follow-up than those in the expressive version (M=4.0, Mdn=5.0, SD=2.0, SE=42, see Figure 24).

Reactions to warning 1: do not enter room 5 Most participants followed up on the warning not to enter room 5 anymore sent to them. 33 of the 44 participants did indeed not enter the room after they had read the message. To see whether there was a difference between the conditions on whether participants entered the room, or not, after the warning a Chi-square analysis was carried out. Its results indicate that there was a significant association of whether a participants received an expressive or non-expressive message to no longer enter room 5 (X²(1)=5.132, p(2-sided)=.026, see Figure 25). 9 out of 23 (39%) participants in the expressive condition entered
the room after the warning, while only 2 out of 21 participants (9.5%) did so in the non-expressive condition. Based on the ‘odds ratio’, the odds of a participant entering room 5 after the warning was in the expressive version was 5.5 times higher than in the non-expressive condition.

![Graph showing effects of expressiveness on willingness to follow-up system advice. 95% CI.](image)

**Figure 24:** Effects of expressiveness on willingness to follow-up system advice. 95% CI.

![Graph showing number of participants entering room 5 after system warning in expressive and non-expressive conditions. 95% Confidence Interval error bars.](image)

**Figure 25:** Number of participants entering room 5 after system warning in expressive and non-expressive conditions. 95% Confidence Interval error bars.

Interestingly, while this difference was apparent in participants’ behaviour, this difference was not reflected in responses the participants themselves provided to the system. Of the 11 participants who entered room 5 after the warning, only one actually pushed the ignore button instead of the OK button. Another 2 participants did push the ignore button, but did not actually enter the room after the warning. Even though the perceived difference between pushing the OK and ignore button in this case might not have been very great to the participants, this does illustrate that participants’ responses to autonomous systems might not reflect their actual behaviour.
7.4 RESULTS

Reactions to warning 2: leave the building

The difference found for the first warning above, was not found for the warning to leave the building. Most participants did decide to look for additional containers, instead of immediately leaving the room. Overall, participants stayed an average of 52 seconds after reading the warning (SD=68, SE=10). Participants in the expressive condition did remain longer in ‘the building’ than those in the non-expressive condition, remaining in ‘the building’ for an average of 69 seconds (Mdn=23, SD=79, SE=17), compared with 34 seconds in the non-expressive condition (Mdn=18, SD=47, SE=10). This difference however was not significant (U=194.500, p(2-tailed)=.269).

The differences outlined above indicate that the non-expressive messages promoted more follow-up behaviour than the expressive messages did for both system requests and warnings.

7.4.4 Effects of social expressiveness on trust

We expected the expressive version of the system to be perceived as more caring and to be trusted more by the participants. The analysis on dependability, source credibility and trust in the information show a different result.

![Figure 26: Mean scores on ‘I trusted the actions of the system’ in the expressive and non-expressive condition. 95% CI.](image)

7.4.4.1 Effects on dependability

There was no significant difference between the conditions on the perceived dependability of the system (t(40)=1.085, p(1-tailed)=.14). Participants in the non-expressive condition rated its dependability as slightly higher, with an average of 5.0 (SD=1.2, SE=.27) compared with an average of 4.6 (SD=1.1, SE=.24) in the expressive condition. When further investigating this result, a significant difference was found on the item ‘I trusted the actions of the system’. Participants in the non-expressive condition reported a higher trust in the system’s actions (M=5.5, SD=1.3, SE=.29)
than participants in the expressive condition (M=4.5, SD=1.4, SE=.29; t(41)=2.324, p(2-tailed)=.025, Figure 26).

7.4.4.2 Effects on source credibility

No significant difference was found between the conditions on the perceived source credibility (t(35)=.307, p(1-tailed)=.15). Participants in the non-expressive condition rated its credibility at an average of 4.7 (SD=.71, SE=.17) compared with an average of 4.4 (SD=1.0, SE=.23) in the expressive condition. When further investigating this result however, we did find a significant difference on two specific items from the scale. Participants rated the non-expressive version as significantly more selfish (M=5.7, SD=1.4, SE=.31) than those in the expressive version (M=3.9, SD=1.7, SE=.36; t(41)=3.761, p(1-tailed)=.001, scale ranging from 1=unselfish to 7=selfish, Figure 27a). Surprisingly, they also rated the non-expressive version as warmer (M=4.8, SD=1.5, SE=.33) than the expressive version (M=3.7, SD=1.6, SE=.34, t(40)=-2.158, p(2-tailed)=.037, scale ranging from 1=cold to 7=warm, Figure 27b).

![Figure 27: Perceived selfishness and warmth of the system in expressive and non-expressive condition. 95% CI](image)

7.4.4.3 Effects on trust in information

A significant difference was found between the conditions on participants trust in the information provided by the system (t(41)=1.833, p(1-tailed)=.037, Figure 28). However, the information was trusted more when coming from the non-expressive system (M=5.1, SD=1.0, SE=.22) than when coming from the expressive system (M=4.5, SD=1.1, SE=.24).
7.4 Results

7.4.5 Effects of social expressiveness on participants’ perception of system goals

Most participants perceived the system as caring most about their well-being (N=28, 64%). Participants’ answers on their perception of the system’s main goal on the questionnaire were grouped in either caring about the participant’s wellbeing or caring about the building, machines and containers (Table 20). We expected that the more expressive version which expressed empathy to the participants’ situation would result in more participants perceiving the system as mostly caring about them. While a Chi-square analysis did not reach a .05 level of significance in two sided analysis (X²(1)=5.69, p(2-sided)=.062, the proportion of participants in the expressive condition who perceived the system as more caring about them (11 of 19, 58%) was actually lower than in the non-expressive condition (17 of 19, 90%). We suggest that further research is necessary to assess the effects of social expressiveness on perceptions of a system’s intentions.

7.4.6 Effects on collaboration/working relationship between the participant and system

7.4.6.1 Effects on perceived closeness

There did not appear to be an effect of the system’s expressiveness on participants’ perception on how close their working relationship with the system was (U=184.500, p(2-tailed)=.240). Participants scored the relationship as an average of 3.8 (Mdn=4.0, SD=1.0, SE=.23) on the 1-6 pictorial closeness scale in the non-expressive condition. Participants in the expressive condition rated the relationship with an average 3.4 (Mdn=4.0, SD=.99, SE=.21).

Figure 28: Effect expressiveness on trust in provided information. 95% CI.
**Table 20: Perceived goal of system, all participants**

<table>
<thead>
<tr>
<th>Goal chosen</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Care for participant</td>
<td></td>
</tr>
<tr>
<td>'Keeping me safe'</td>
<td>N=28, 64%</td>
</tr>
<tr>
<td>Care for artefacts</td>
<td></td>
</tr>
<tr>
<td>'Preserving the building'</td>
<td></td>
</tr>
<tr>
<td>'Locating the chemical containers'</td>
<td>N=4, .9%</td>
</tr>
<tr>
<td>'Preserving the machines in the machine room'</td>
<td>N=2, 4% (reclassified from open answers)</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Open text e.g. 'both keeping me safe and preserving the machines', 'The system doesn’t care more or less about anything'</td>
<td>N=4, .9%</td>
</tr>
</tbody>
</table>

7.4.6.2 *Effects on perceived control*

No significant difference was found between the conditions on whether participants considered themselves as being in control or the system being in control (t(42)=.565, p(2-tailed)=.575). In both conditions participants scored around neutral on whether they or the system were in control more (Me=3.4, SDe=1.5, SEe=.32; Mne=3.6, Sdne=1.7, SEne=.36).

7.4.6.3 *Effects on perceived disruptiveness*

Perceived disruptiveness was also not significantly affected by the system’s expressiveness (U=211.000, p(2-tailed)=.614). Participants in the non-expressive condition rated the system with an average 4.6 on a 1-7 scale (Mdn=5.0, SD=1.5, SE=.33), in the expressive condition it was rated with an average 4.9 (Mdn=5.0, SD=1.5, SE=.32).

7.4.7 *Effects of social expressiveness on participants’ emotional experience*

Participants rated their emotional experience as a relatively positive one (M=3.8, SD=.39, range=3.0-4.5 on a scale from 1-5, 5 being most positive). No significant difference was found between the conditions (t(38)=1.266, p(2-tailed)=.21). Participants in the non-expressive condition reported an average of 3.9 (SD=.41, SE=.093); participants in the expressive condition an average 3.8 (SD=.37, SE=.081). Overall, participants did not appear to experience much stress during the experiment. On the stress and anxiety subscale (stressed, nervous, afraid, scared) scores ranged...
from 1-3.25 on the scale of 1 (very slightly, or not at all) to 5 (extremely), with an average score of 1.5 (SD=.54) with no significant difference between the conditions (U=177,000, p(2-tailed)=.122). It appears that the socially expressive statements, meant to express empathy to the users’ experience during the experiment did not affect their emotional experience overall. However, when further exploring the negative emotion items, interesting significant differences were found on the items alert and ashamed.

The overall score on ashamed was very low (M=1.1, SD=.39, SE=.06 on a scale of 1-5). However, participants in the expressive condition expressed being more ashamed (U=198,000, p(2-tailed)=.043; M=1.23, Mdn=1.0, SD=.53, SE=.113) than those in the non-expressive condition (M=1.0; Mdn=1.0, SD=.00, SE=.00). On further inspection only 4 participants reported a score higher than 1 on the ‘ashamed’ item; all of these four participated in the expressive condition. This might indicate that socially expressive system statements might come across as somewhat embarrassing for a small selection of users.

A similar significant difference was found on how alert participants had rated themselves (U=160,500, p(2-tailed)=.041, see Figure 29). Participants in the expressive condition rated themselves as less alert (M=3.7, Mdn=4.0, SD=.78, SE=.16) than those in the non-expressive condition (M=4.1, Mdn=4.0, SD=.73, SE=.16). Interestingly, one of the expressive statements of the system included the phrase ‘you appear quite alert’ for one of the found containers. This indicates that this type of positive statement, meant to be empathic to the participants’ success in finding another container, might not necessarily yield positive results.

![Figure 29: Effect expressiveness on reported alertness. 95% CI.](image)

**7.4.8 Effects of participant personality**

Empathy and extraversion were taken into account as possible personality traits affecting interaction and attitudes; effects of participants’ own empathy levels are discussed first, extraversion is discussed afterwards.
7.4.8.1 Participant personality: empathy

To investigate the effect of empathy as a personality characteristic on participants’ reaction to the system, participants were divided into two groups at the median score (Mdn=3.1) on the empathy scale: one group scoring below the median (relatively low in empathy, N=21) and one group scoring above the median (relatively high in empathy, N=20). The found significant effects are discussed below.

Interaction effect participant empathy and system expressiveness on emotional experience

Two-way ANOVAs yielded interaction effects between empathy as a personality trait of the participants and social expressiveness of the system on the reported emotional experience. A significant interaction effect was found between expressiveness of the system and participants’ empathy score (F(1, 38)=4.330, p=.045, see Figure 30).

Simple effects analysis showed that in the non-expressive condition low empathy participants had a more positive emotional experience (M=4.1, SD=2.9) than high empathy participants (M=3.6, SD=37). This difference was absent in the expressive condition (Mhigh=3.8, SDbhigh=.49 vs. Mlow=3.7, SDblow=30). As Figure 30 shows, participants scoring low on empathy reported a less positive experience when using the expressive system than when using the non-expressive system; this difference was significant (F(1, 35)=9.56, p=.004). The figure also shows that participants scoring high on empathy in contrast reported a slightly more positive emotional experience in the socially expressive condition than in the non-socially expressive condition. This difference however was not significant (F(1, 35)=.03, p=.86). This indicates that high and low empathic people react differently to semi-autonomous systems that show socially expressive behaviour. No interaction effect was found for the specific emotion items ashamed and alert on which the social expressiveness condition had a direct effect.

An interaction effect was also found for gender and expressiveness of the system on reported emotional experience, with male participants having a more positive experience with a non-expressive system overall (F(1, 36)=7.51, p=.010) and having a less positive experience in the expressive condition than female participants (F(1, 36)=5.46, p=.025). This however can be explained by female participants scoring higher on the empathy personality scale (M=3.4, SD=.67) than male participants (M=2.9, SD=.67; t(39)=2.544, p(2-tailed)=.015).

Direct effect participant empathy on follow-up warnings

Empathy as a personality trait appeared to result in a difference on follow-up on the system warning to not enter room 5 anymore. Two out of the 21 (9.5%) participants scoring low on empathy entered the room after the warning, whereas 7 out of 20 (35%) participants scoring high on empathy did enter the room. High empathy participants thus appeared more likely to enter the room after the warning than low empathy participants. This particular difference was not significant at a .05 level in this study (X²(1)=3.881, p(2-sided)=.067). When further investigating, participants scoring high on empathic personality traits group did spend significantly (U=140.000, p(2-tailed)=.022) longer in room 5 after the system’s warning (M=15, Mdn=0, SD=21, SE=5) than participants who scored low on empathy (M=3, Mdn=0, SD=10, SE=2). Further investigation of possible effects of personality on compliance would be useful.
7.4.8.2 Effects participant personality: extraversion

Participants were also divided into two groups at the median score (Mdn=3.25) on the extraversion scale: one group scoring below the median (relatively low in extraversion, N=21) and one group scoring above the median (relatively high in extraversion, N=21). No effects of expressiveness were found on any of our dependent variables. More extravert participants did not appear to react different to the system and its expressiveness than relatively introvert participants.

7.4.9 Correlations for trust and participants’ emotional experience

Correlations were calculated between the main variables (Figure 31) to further explore the relationships between user trust, perceived empathy and the user’s emotional experience. Pearson’s correlation coefficient was used for parametric variables. Spearman’s rho was used for non-parametric constructs. All correlations reported below are significant at the p(1-tailed)<.05 level.

Perceived empathy of the system has a significant relationship with perceived dependability (r=.36) and perceived source credibility of the system (r=.50). These findings confirm the expected relationships between variables related to socially expressive behaviour; they suggest that perceived empathy relates to trust in a system. Perceived empathy however is not directly related to trust in the information provided; this may suggest that while users might have confidence in the system itself and its credibility, it is not directly related to trust in the specific information delivered in an empathic manner. This also illustrates the need to distinguish
between trust in a system as a whole and trust in its individual decisions, requests and recommendations (cf. Chapter 3, Cramer et al., 2008).

Trust in the information, dependability and source credibility are related to users’ willingness to follow-up on the system’s messages. The amount of time spent in dangerous areas was inversely related to trust in the information provided (rho=-.35, p(1-tailed)=.10) and willingness to follow-up on the system’s advice (rho=-.53, p(1-tailed)<.001). The number of system information requests that the participants answered was related to dependability of the system (rho=.35, p(1-tailed)=.015), source credibility (rho=.38, p(1-tailed)=.012), willingness to follow-up on these requests (rho=.48, p(1-tailed)=.001) but also willingness to follow-up on the system’s warnings (rho=.42, p(1-tailed)=.003). The number of requests answered was not directly related to trust in the information provided by the system.

The number of answered requests was however directly related to participants’ emotional experience (rho=.29, p(1-tailed)=.040). Participants’ emotional experience was also significantly related to perceived dependability, source credibility and trust in the information provided by the system and willingness to follow-up on the system’s advice (Figure 32). These results indicate that user emotions play a role in interaction with semi-autonomous systems and influencing this experience could potentially influence trust in a system and its information.

**Figure 31**: Main significant correlations, Pearson’s r / Spearman’s rho. All p(1-tailed) < .05. Correlations for emotional experience are left out for clarity and shown in the next figure.
The found correlations for perceived empathy with source credibility and system dependability indicate that appropriate empathy with the user’s affective experience can have an important impact. This is consistent with the previous research as discussed in section 7.2. However, this study shows that socially expressive, empathic systems, in contrast to hypothesis 1, will not necessarily be trusted more. Instead, in this study’s urgent context the non-expressive system’s actions and information it provided were trusted more.

Hypothesis 3 suggested that empathic expressiveness would have a positive effect on the willingness to reply to system requests. This was not the case. Our expectation that warnings or urgent advice from a system posed in a more expressive manner would be followed up on less was confirmed. This indicates that the empathic advice might have been too friendly and was taken less seriously. The manner in which a message is delivered has to be congruent with its content and context. This appears in line with findings of Goetz and Kiesler (2002) who found that while people enjoyed interacting with humorous robots, they complied more with more serious robots. Clear instructions on what to do, might be more persuasive in some cases (especially in urgent contexts) than socially expressive, empathic behaviour, especially as time-constraints increase the importance of (perceived) efficiency. Users’ liking of a system might in some cases have to be sacrificed in order to yield better results on compliance.

Some factors in the experiment set-up could have played a role in for example compliance to the system’s requests. As Eckles et al. (2009) point out, embodiment, familiarity and perceived proxemics can affect social reactions. The autonomous system in this study was not embodied; instead the monitoring system interacted via textual messages in a relatively unfamiliar format. A number of participants during their interviews mentioned that the socially expressive version’s messages would perhaps be more appropriate if they came from a human being and were not pre-
programmed. Empathic interaction might also require time to build a relationship, both to learn which responses are appreciated by users and to increase acceptability as a ‘stranger’ commenting on your emotional state might be disconcerting and uncomfortable. The specific system or agent that is showing social behaviour will affect people’s responses. A combination of expressive messages with other social, more subtle, non-verbal cues could yield different results.

The urgent context of this experiment might also have led to more instances of social expressive behaviour being perceived as inappropriate as they may lead to the communication with the system being perceived as less efficient. It’s also worthwhile considering that the stress levels in urgent situations could have an influence on how effective socially expressive messages are. When deciding to implement social behaviours and explicit statements on the user’s affective experience, contextual evaluation will be crucial.

Socially expressive, empathic behaviour also did not yield positive effects for disruptiveness, closeness of the relationship and the level of control participants thought they had over the situation (contrary to hypothesis 2). Even though the system was empathic to task success and challenges to this success in the expressive version, this did not appear to change participants’ perceived or actual performance (contrary to hypothesis 4). Social expressive behaviour was expected to have a positive effect on users’ emotional experience, but this effect was not observed in this study. Instead, participants indicated feeling less alert when interacting with the socially expressive system. A small number of participants even appeared to find interacting with such a system slightly embarrassing. These more negative responses in the empathic condition were unexpected, but consistent with the findings of Prendinger et al. (2006), who found that a positive empathic character in a competitive game context was experienced to be more stressful and irritating than a negative empathic agent. In the context of the game in Prendinger et al. (2006) participants did not expect the positive behaviour of the agent and this incongruence may have caused their negative response; a similar effect might have occurred in this study.

This study confirmed that personality can affect reactions to social behaviour and users’ affective experience (hypothesis 5). Matching user and system personality appears beneficial regarding empathy. Low empathy participants reported a significantly less positive affective experience (on the scale consisting of items such as excited, upset, and alert) when using the socially expressive, empathic system. It appears low empathy participants could be less forgiving of ‘inappropriate’ expressive statements. We also found indications that users’ own empathy level can directly affect compliance. More empathic participants took longer to comply with warnings. A possible explanation would be that participants high in empathy may care more about their task performance. Whether this is actually the case and whether such behaviour will also be shown in other situations remains to be seen. We did not find any effects of extraversion. Further analysis of possible effects of personality would also be very useful, especially since current information on the effects of personality on interaction with autonomous systems is fragmented as most studies do not take into account the same personality factors.
Some explanations can be offered for these findings. The socially expressive, empathic condition was rated as more empathic on the manipulation check in the questionnaire. However, the expressive dialogue in hindsight was less than ideal. From the interviews it became clear that what appeared socially expressive and motivating to some participants, was considered socially awkward and inappropriate by others. It appears in some cases participants in the expressive condition recognized that the system was attempting to be empathic to their emotions, but that they were skeptical about the system’s capabilities in this regard. Socially expressive and empathic system reactions that might not fully capture the user’s actual affective state, or are seen as exaggerating the user’s experience, might do more harm than good to the user’s trust in a system.

When actively reflecting on the messages, opinions on the socially expressive condition described during the interviews ranged from ‘encouraging’, ‘compassionate’ to ‘inappropriate for the situation’. One participant even mentioned that he thought the messages were so encouraging and praising that they appeared ‘sarcastic’. This suggests that perceptions of a system’s intentions could be affected by its social behaviour. In this study, the proportion of participants in the expressive condition who perceived the system as more caring about them was not significantly lower than in the non-expressive condition. However, coupled with the interview statements above, these results do suggest a potential doubt in participants’ mind over the sincerity of expressive system statements. This could seriously diminish willingness of users to respond to a request, especially when the social behaviour is seen as inappropriate. In worst cases social expressiveness coupled with perceived differences between the system’s goal and those of the user could result in feelings of being manipulated by a system.

### 7.5.1 Additional implications for designing context-aware agents that request and provide information

In addition to the study’s implications on implementing social system behaviour discussed above, observing the 50 participants while they interacted with the system and responded to its requests and warnings offered insight in a number of problems that can lead to guidelines for design of autonomous systems.

#### 7.5.1.1 Ensure an appropriate level of user understanding

Even some of the highly educated participants reported during the interview that they did not know ‘how intelligent systems are nowadays’. In case of hazard monitoring systems this might result in users taking risks on the assumption that the system has a full overview of their surroundings or instead dismissal of crucial system warnings when the system’s capabilities are underestimated.

Ensuring a full understanding is especially crucial when requesting users to supply information. The system’s requests in this study deliberately were somewhat ambiguous. This led to a number of issues reported in the interviews: participants reported wondering which level of detail was necessary and having difficulty
deciding which information from their surroundings was relevant. They reported wondering to which detail they had to report information, how the system would process their input and whether they used the right terms the system could indeed use. Making clear why a system needs more additional information, how it reaches its conclusions and possible limitations to its advice is crucial.

7.5.1.2 Don’t ‘overtrust’ user-provided information

Related to the issue of understanding above, a number of problems with participants’ input stood out in observation. Some participants reported being curious about what the system would answer and tried non-sensical answers to see how it would respond, even though they did realise they were being observed and were competing for points. Other participants only partially took into account the instructions given or misunderstood instructions. Participants also were inconsistent in their reporting, for example in regards to the terms they used in different messages. Verification with multiple sources or confirming information from users themselves (at a suitable moment) is very important.

7.5.1.3 Consider that the outside world will interrupt system requests

Even in this relatively clear lab set-up, participants often appeared to have difficulty deciding whether to follow-up on requests and advice. This meant they lost valuable time. Even if they had decided to follow-up on a message, their surroundings and task were in cases more salient than the messages. Participants were distracted even while they had already started to follow up on the system’s requests. Systems need to take these distractions into account by for example supporting recovery of earlier messages and providing reminders, while also making sure users do not repond to messages that have been rendered irrelevant due to changes in the current situation.

7.5.1.4 Consider user, context and setting - including cultural differences

This study shows that social behaviour will have to be adapted to user personality, context and content of the information provided. Social expressive behaviours are also culturally-dependent. One participant in this study for example mentioned ‘that sort of American stuff’ wasn’t for him. The extent to which a system will come across as polite versus insincere and manipulative depends on careful and considerate design. Evaluating of social behaviours has to be carried out in multiple settings before a system’s social dialogue can be considered to be successful. It has to be taken into account that empathic ‘lip-service’ to the user’s experience is not enough. Taking into account the user’s experience should not only consist of socially expressive statements referring to the user’s affective state; anticipating whether a user can be interrupted for example and acting accordingly is just as important.
7.6 CONCLUSION

This study shows that even though the literature shows great potential for social expressive systems that are empathic to the user’s experience, inappropriate socially expressive behaviour can have less desirable effects.

The correlations found in this study indicate that perceived empathy can indeed increase trust in a system, however, the comparison of the two conditions yielded a different result. Even though the socially expressive, empathic condition was recognised as more empathic, it did not increase trust in the system and did not increase compliance with its requests and information it provided. Instead, the study’s findings indicate that (inappropriate) social, empathic expressivity can change perceived systems’ intentions and lead to less compliance with a system’s warnings. Perceived empathy was correlated to perceived dependability of a system and its source credibility, but not directly to trust in the provided information. We have to clearly distinguish between effects on trust in a system overall and in its specific advice and requests. 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.

Clearly this study has its limitations as a lab study in which most participants actively reflected on their interaction after the experiment session. Arguably, even while the expressive version was rated as more empathic in the survey, the expressive statements in hindsight were less than ideal for many participants. During the interview, some participants reported being encouraged by the expressive system’s enthusiasm, others found it inappropriate for the situation, or culturally awkward or insincere. As this study illustrates, it is important to evaluate the effects of social behaviour and not just investigate whether (an attempt at) social behaviour is recognised. In this study, the system’s attempts at empathy resulted in counterproductive effects. Simply making messages appear ‘friendlier’ is not enough in order for empathic system behaviour being experienced as a positive interaction feature. The social expressiveness and level of empathy need to be congruent to the context, individual user characteristics and the message delivered by the system. The source of the behaviour will play a role; is the behaviour shown by a distant system, a service, a local, personal application or a distant human organisation? Careful consideration of which behaviour will fit the context, user and the system’s purposes is crucial.

It appears that a considerable portion of the participants in the socially expressive condition recognised the system was trying to take their emotions into account, but that they considered its statements as inappropriate. Letting users know that the system (or its designers) realise they have feelings, may raise expectations; failing to live up to these expectations and not reacting in a socially acceptable manner can hurt their experience. Presenting users with recognisable, but inappropriate attempts at mimicking social behaviours or addressing user affect in (blunt) explicit statements does not constitute empathic interaction, but rather just may be seen as
‘trying too hard’. It might be helpful for HCI researchers interested in, for example, empathy to learn from researchers in other fields. Pedersen (2009) for example provides a methodological review of empathy studies in the context of medicine. He points out that a shared view on what is meant by empathy and how it can be studied is absent; a wide variety of definitions and measurement methods are used. He also warns against taking an overly quantitative approach to measuring empathy and emotions that neglects the concrete feelings and richer experiences of the people involved. Future studies need to take these factors into account and add to the growing body of social expressivity and empathy research to provide a fuller picture. It would be beneficial to tackle these issues for HCI research on empathy towards users now, as HCI research on such social behaviours is still in its relatively early stages. Social expressivity has great potential, but if the displayed behaviour appears inappropriate or, in worst cases, manipulative, both system acceptance and the user’s experience will suffer.

ACKNOWLEDGEMENT

This research is partly funded by the EC FP7-ICT project DIADEM, ref. no: 224318.