Simulation of an Emergency Situation for the Purpose of Evaluating a Crowd-Based Rescue System

Svenja Polst
(s1228765)

Faculty of Behavioural, Management, and Social Sciences,
University of Twente,
Enschede, the Netherlands

ln partial fulfillment of the master's thesis requirements
Primary supervisor: Prof. dr. Jan Maarten Schraagen
Secondary supervisor: Dr. Martin Schmettow
External supervisor: Eduard C. Groen

22-04-2016

Acknowledgement

I like to thank Jan Maarten Schraagen, Martin Schmettow, and Eddy Groen for your supervision, the feedback and your time you put in reading the different versions of the thesis.

Furthermore, I like to thank Fraunhofer IESE for offering the opportunity to write my thesis there and the project leaders of RESCUER, Karina Villela and Claudia Nass, for their support.

I thank Risa and Alberto for their help, especially at the day of the experiment, and all the other students at Fraunhofer IESE who supported the experiment by participating in the (pilot-) study or by being one of the assistants during the experiment.

Abstract

In the context of the RESCUER Project, the need for a simulation of an emergency arose that is suitable to evaluate the RESCUER App, a crowd-based rescue system. To the best of our knowledge, such a simulation has not yet been described in the literature. Therefore, a simulation was created that takes the use context of a crowd-based rescue system into consideration and that induces stress. Literature, mainly about laboratory studies, was reviewed to gain knowledge about stressors. Stressors were selected based on their suitability for the simulation. Cognitive, acoustic, and emotional stressors were selected as well as a feeling of uncontrollability and unpredictability were created. Additionally, assistants and confederates were included to evoke stress. The simulation was tested with the RESCUER App. The stress manipulation was evaluated by analysing saliva cortisol and subjective ratings (stress rating, SAM) with the group (control or experimental), the moment of measure (before, during, after), and app use (participants were provided with the app or not) as predictors. The context was evaluated by observing the participants and by analysing other data obtained from the participants. The results revealed that the simulation was successful for the experimental group, even though the cortisol measures did not show a stress reaction. The control group indicated through subjective ratings more stress than expected. However, those ratings were in conflict with other results, hence the ratings were probably due to a testing effect. Therefore, the simulation for the control group has to be improved. Future research should address whether the simulation is also suitable to evaluate a crowd-based rescue system from the perspective of first responders, that means while moving from a safe place to the incident.

Contents

Glossary

The definition of crowd is derived from http://www.oxforddictionaries.com. The definitions of emergency, eyewitness and incident are copied from Nass et al., 2015.

Crowd: a) A large number of people gathered together in a disorganized or unruly way
b) An audience, especially one at a sporting event

Emergency: An event or situation which threatens to impose serious damage to human welfare (e.g., loss of human life, illness, injury, homelessness, damage of property, disruption of a supply of money, food, water, systems of communication, facilities for transport, or services relating to health) or the environment (e.g., contamination of land, water or air with biological, chemical or radioactive matter, or disruption or destruction of plant life or animal life).

Eyewitnesses: People at the place of the incident that caused the emergency situation.

Incident: Can cause damages to life and property and seriously affect the image of a business and/or a country. The concept of incident as defined here includes the concept of accident.

RESCUER App: An app that is developed in the scope of the RESCUER Project. The app has the objective that a crowd can efficiently provide the command centre of rescue services with information about an incident.

Abbreviations

RESCUER: Reliable and smart crowdsourcing solution for emergency and crisis management

1. Introduction

Current rescue systems do not provide all the functions required to organize a large-scale rescue operation. The functions meant here are alternative communication channels in case the mobile network is overloaded, integration of the communication systems used by police officers, fire-fighters, and other rescue workers and integration of information given by a crowd. A consortium of companies and research institutions from the European Union and Brazil was formed to build a rescue system that integrates the aforementioned functions. The rescue system is built in a project partially funded by the European Union called 'Reliable and smart crowdsourcing solution for emergency and crisis management', abbreviated with the acronym RESCUER (Nass et al., 2015). Its mission statement is "*The RESCUER project aims at developing a smart and interoperable computer-based solution for supporting emergency and crisis management, with a special focus on incidents in industrial areas and on large-scale events.*" (RESCUER, 2016).

Within the project, the RESCUER App was developed, which enables the crowd to send information about an incident directly to the command centre. The design of the app takes into consideration that stress, which is evoked by an emergency situation, impairs cognitive processes (Nass et al., 2015). Three screens for interaction with the command centre were designed that differ in their cognitive demand (see figure 1). The first screen is meant to be used close by the incident and the third one when the user is in a safe place. To ensure an effective use of the app, it has to be evaluated in an emergency situation or a simulation of such a situation. So far, the app had been evaluated twice, once by attendees of an event in soccer stadiums and by people in industrial parks, but without inducing stress, (Nakagawa, Soares Santos, Bueno Ruas de Oliveira, & Duran, 2015) and the second time in a laboratory setting with a stress manipulation (Nass et al., 2015). Stress was induced by an auditory n-back task (Nass et al., 2015). In both evaluations, the participants had to complete some pre-defined tasks, such as 'Describe the properties of the fire' while standing or sitting still, thus, without simulating flight behaviour. Flight behaviour should be simulated since it could impair the app use (Delignières, Brisswalter, & Legros, 1994). Stress has to be induced since it affects several stages of information processing (Duncko, Johnson, Merikangas, & Grillon, 2009; Ozel, 2001). A third evaluation needs to be conducted which simulates the use context, an emergency situation, as realistically as possible. Without another evaluation an effective use of the app in a real emergency is not ensured. Users may not be able to use the app under stress, resulting in providing wrong information or no information at all.

To the best of our knowledge, there is no simulation of an emergency described in the literature which resembles the use context of a crowd-based rescue app, such as the RESCUER App. The current literature about evaluating mobile applications for emergency situations is very limited. There is a study about a system for rescue workers to communicate with the command centre by sending videos, which was tested in real emergencies (Bergstrand & Landgren, 2009). Additionally, there are proposals for how to test an app for professional rescue workers, such as police officers, but those

proposals do not take stress into consideration (Streefkerk, 2011). Therefore, we created a simulation of an emergency situation ourselves that is applicable for evaluating a crowd-based rescue system, like the RESUCER App. The simulation consists of a selection of stressors and the creation of the use context. The suitability of the simulation was tested by evaluating the RESCUER App. Therefore, we will subsequently describe the context of an emergency in which a crowd is involved since the context defines the requirements for the simulation. Thereafter, stress, stressors, ways of measuring, and moderator variables will be described so that the reader may be able to reconstruct our decisions concerning the simulation. The simulation of an emergency situation is subsequently described. The simulation of the use context could be adapted to the rescue system. In this study, it is adapted to the evaluation of the RESCUER App.

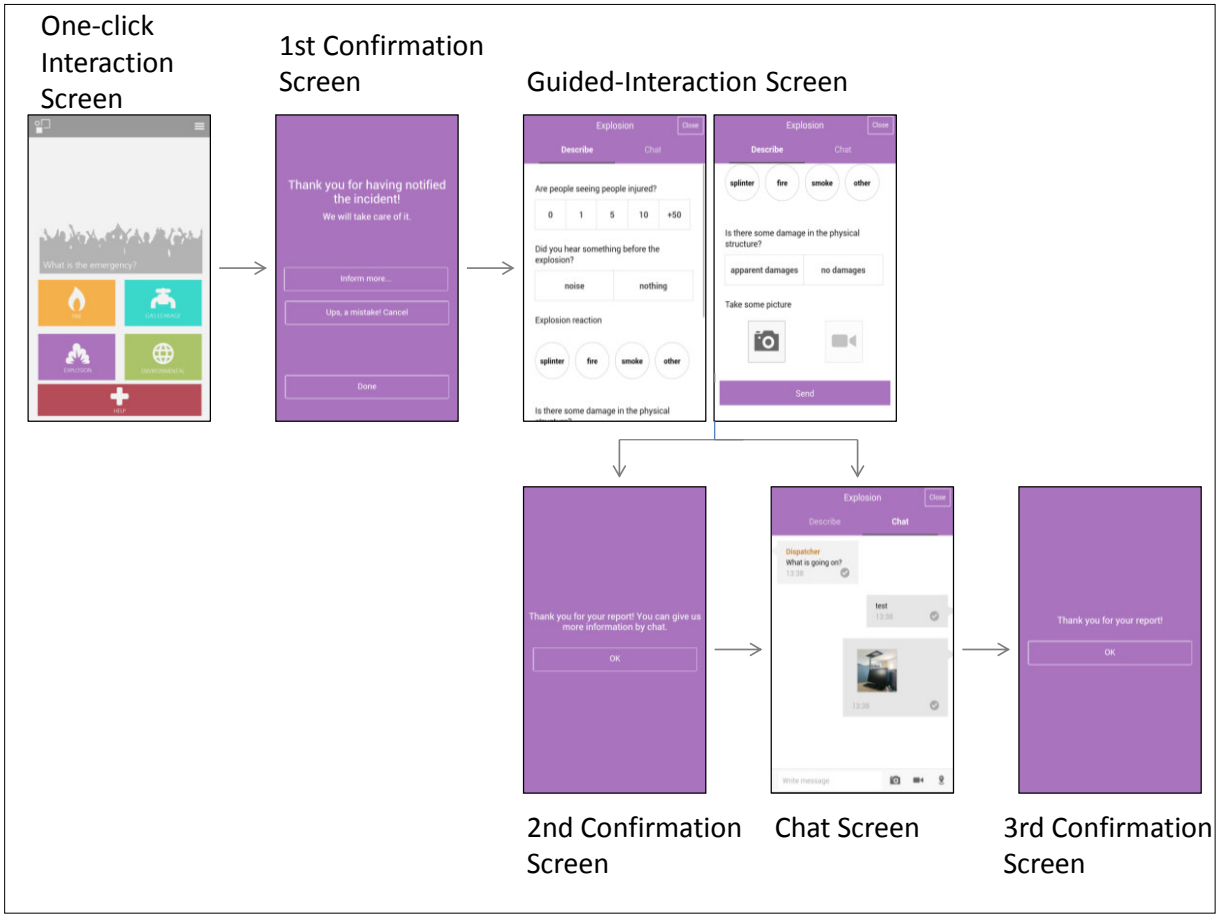


Figure 1. Illustration of the screens of the RESCUER App and their order for reporting an explosion. At the One-click interaction Screen, the user can send information about the emergency type with a single click. The confirmation screen informs about the information sent. The Guided-Interaction Screeen asks more specific questions about the reported emergency type. The user can directly continue to the chat screen or send the answers to the questions. After sending, another confirmation screen appears. The Chat Screen allows the user to send text messages to the command centre of the rescue services. The last screen confirms the sending of information through the chat screen.

1.1. Use Context of a Crowd-Based Rescue App

A crowd is a large group of people who gather, for instance, at public places, at large-scale events or in industrial parks. The crowd could consist of people from various cultures. Crowd-based rescue apps are likely to be used in an emergency by eyewitnesses and first responders. In case of an emergency, those people have to cover quite a distance to reach a safe area. They show flight behaviour, which is reflected probably in walking fast or even running. However, there might be emergencies that do not require that people run or walk fast. On the way to safety, people pass through a maximum of three zones depending on where they are respectively the incident. The zones are called the hot zone, warm zone, and cold zone. In the hot zone, individuals are close to the incident. The most cognitive resources are needed to flee from the incident (Nass et al., 2015). In the warm zone, individuals have already made some physical and timely distance from the incident thus, the immediate danger is lower and fewer cognitive resources are demanded for fleeing and the experienced stress is lower (Nass et al., 2015). Cold zone means the individuals left the area at risk and are able to calm down. Thus, the available cognitive resources and experienced stress of the zones differs. On the way to the cold zone, it is likely that people enter unknown (event) areas or unfamiliar parts of an industrial park. In all areas, individuals probably come across environmental cues about the incident, such as fire, smoke, damaged objects, and injured persons. Not everyone in the crowd would use an emergency app, since it requires owning a smartphone, which is charged sufficiently, and the belief in the benefits of the app. The requirements for the simulation are summarized in table 1.

1.2. Stress

Three common stress approaches, the stimulus-based, the response-based, and the transactional stress approach, are introduced to help the reader understand topics that are discussed later on.

The stimulus-based stress approach emphasises that a stimulus that causes psychological distress or physiological impairment or deterioration is a stressor (Hobfoll, 1989). A stressor can be a condition such as heat or noise (Sanders, 1983). The human reaction to these stressors is called stress. Stressors can be categorized as acute and time-limited, such as a visit to a dentist, stressor sequences such as losing a job, chronic and intermittent stressors such as writing exams or chronic stressors such as chronic illnesses (Hobfoll, 1989, based on Elliot and Eisdorfer, 1982). An emergency situation is an acute and time-limited stressor. The stimulus-based approach is criticized for not taking individual differences into consideration (Hobfoll, 1989).

The response-based approach views stress from an endogenous perspective. The stress definition by Hans Selye is: “Stress is the nonspecific response of the body to any demand (Selye, 1976, p. 53). Selye called the physiological reaction to stressors the ‘general adaption syndrome’ (Kalat, 2008). At the first stage of this syndrome, called ‘alarm’, the activity of the sympathetic nervous system increases (Kalat, 2008). This stage prepares the body for a fight-or-flight reaction.

Table 1

Overview of the Requirements for the Simulation of the Use Context of a Crowd-Based Rescue System

Category	*Has to be simulated/ taken into account*	*Explanation/ Example*
Humans	Crowd	A large group of people
	Users/ Non- users	Not everyone uses the rescue system The users could be eye-witnesses or first responders
	Culture	Various cultures could be represented in the crowd. Culture may influence how people react to an emergency, e.g. due to how the culture deals with uncertainty
	Flight behaviour	Running or walking fast to leave hot or warm zone
Location	Large area	E.g. : Event areas, industrial areas, public places, such as malls The distance to a safe area could be large
	(partly) Unknown location	It is likely that several people pass through areas on the way to safety which are unfamiliar to them if they are not going exactly the way they entered
Emergency	Incident	To trigger flight behaviour, people have to be aware that an incident had occurred, either by perceiving cues directly related to the incident, e.g. fire or smoke, or by perceiving indirect cues, such as people running away or the sound of a fire alarm
	Three zones	The area around the incident can be divided in three zones Hot zone: Close to incident, imminent danger High stress, few cognitive resources Warm zone: Some distance to incident Moderately stress, moderately cognitive resources available Cold zone: Safe area Low levels of stress, more cognitive resources available
	Environmental cues	e.g. : Smoke, fire, other people who are fleeing
	Stress	Stress and negative emotions are evoked but mostly no panic (Quarantelli, 1975)
Ethics	International ethic standards	The participants have to sign an informed consent and no deception should be applied (see appendix E for more detail)

At the second stage, resistance, the activity of the sympathetic nervous system declines but the Hypothalamus-Pituitary-Adrenal (HPA) Cortex axis is activated. The HPA axis initiates the production of cortisol and other hormones which support a prolonged resistance to stress. These hormones support the body to maintain alertness and to fight illness. The last stage is exhaustion which is characterized by inactivity and tiredness of the body due to exhaustion of the immune system and the nervous system. Kemeny (2003) criticized the general adaption syndrome for assuming a non-specific reaction to stressors. She states that different stressors could elicit different responses depending on whether the threat is evaluated as controllable or uncontrollable. Furthermore, the reaction to stress is said to be influenced by personality as well (Vollrath, 2001). Despite this point of critique, many studies measure stress by analysing the increase of salivary cortisol (e.g. von Dawans, Kirschbaum, & Heinrichs, 2011; Schwabe & Wolf, 2010; Gathmann et al., 2014).

The transactional model states that stress results from an interaction between environment and individual (Hobfoll, 1989). According to this approach, stress results from a mismatch between environmental demand and perceived coping capabilities.

For this thesis, it is not important to apply a single definition since all approaches have some valuable points. The stimulus-based approach helps to understand what stressors are. The response-based approach is the basis for physiological measures of stress, which are explained in 1.2.2. The transactional approach served as the basis for a model that explains the interplay of stress with the time and distance to a threat (Nass et al., 2015) and the influence of the feeling of control on stress.

1.2.1. Stressors

We learnt from the previous section that the simulation is about acute stressors. We have reviewed laboratory studies to learn more about stressors (see appendix A for an overview over those studies).

Stressors can be divided into psychological stressors and physiological stressors (Kolotylova et al., 2008). Physiological stressors are external stressors such as heat, cold and noise (Kolotylova et al., 2008). Psychological stressors can be classified as idiosyncratic or standardized. The idiosyncratic stimuli evoke a higher stress level than standardized stressors since they are adapted to the social-biographic background of an individual. Idiosyncratic stressors are mostly used for evaluating coping strategies in a small number of participants. The standardized stressors were used in the simulation since they are meant for bigger groups of healthy individuals. According to Kolotylova et al. (2008), standardized psychological stressors are emotional stressors, cognitive stressors, and social stressors. For the simulation of an emergency situation, standardized psychological stressors as well as physiological stressors are necessary. In the literature, there is no consistent use of terms related to stressors (e.g. compare Dickerson & Kemeny, 2004; Kolotylova et al., 2008). Within this thesis, we refer to a stimulus or condition that creates the feeling and/or physiological reaction of stress as stressor.

Table 2

Overview over Stressor Categories, the Belonging Stressors, Examples and References

Stressor Category Name	*Description*	*Example*	*References*
Acoustic stressors	Exposure to noise	Random noise, increasing from 78 db(A) to 93db(A)	Kolotylova et al. (2008)
Cognitive stressors	Cause high cognitive load	Mental arithmetic task (e.g. subtracting the number 13 serially starting with 1022)	Kirschbaum, Pirke, & Hellhammer (1993)
		n-back task	Nass et al (2015)
		Stroop task	*Mentioned in* Dickerson & Kemeny, (2004)
Emotional stressors	Inducing negative emotions	Pictures	Lang, Bradley, & Cuthbert (1997), Dan-Glauser & Scherer (2011)
		Videos	Nejtek (2002)
		Words	Bradley & Lang (1999), Schmidtke, Schröder, Jacobs, & Conrad (2014)
		Text	CSEA (2016)
		Sounds	Stevenson & James (2008)
		Odour	Croy, Olgun, & Joraschky (2011)
Motivational stressors	Motivating participants to commit to the experiment	Losing money for each wrong answer	Kolotylova et al. (2008)
Social evaluation	Being judged by others	Simulated job interview	Kirschbaum et al. (1993)
		Singing in front of two confederates	Brouwer & Hogervorst (2014)
		Being watched and video-taped for later analysis	Smeets et al. (2012)
Thermal stressors	Heat and Cold	Putting one hand into 0-4ºC cold water for a maximum of 3 minutes	Schwabe, Haddad, & Schachinger (2008)
Uncontrollability	Not being able to influence an outcome through behaviour	Duration of exposure to a stimuli is randomly determined by a computer	Smeets et al. (2012)
Unpredictability	Receiving only vague or no information	Not knowing duration of a trial	Smeets et al. (2012)

Those stressors can be summarized into categories, which we call stressor categories. Table 2 summarizes stressor categories that are mentioned in the literature about laboratory stress tests.

Some stressor categories are not interesting for the simulation of an emergency situation, because they do not fit an emergency. In an emergency situation, individuals are exposed to acoustic stressors such as noise, for instance, the sound of fire alarm and screaming people. They may see emotional stimuli, for instance, injured persons and crying children, which make them feel desperate. They probably experience the feeling of uncontrollability because they cannot influence the situation through their behaviour, especially when they are close to the incident, and a feeling of unpredictability since they do not know what is about to happen and when the unpleasant situation will be over. Moreover, negative emotions, such as fear, arise. Their cognitive load will be high since they are exposed to many stimuli, they have to find the way to a safe place, and try to make sense of the situation. Thus, social evaluative stressors do not fit an emergency situation. Moreover, motivational stressors are not applicable, too, because adding an external motivator may incite the participants to behave more competitively within a group. Thermal stressors take too much effort to apply in movement and to control. In the following section, the relevant stressor categories are described in more detail.

Uncontrollability and unpredictability. Uncontrollability means that individuals could not influence an outcome through their behaviour (Dickerson & Kemeny, 2004). More precisely, "the presence of continuous or intermittent loud noise, auditory distraction, or other emotionally distressing stimuli without the possibility of a behavioural response" (Dickerson & Kemeny, 2004, p. 361) makes a situation uncontrollable. A characteristic that is similar to uncontrollability is unpredictability. It means, that a subject cannot predict the procedure. Unpredictability is also related to the release of cortisol (Mason, 1968). Several studies do not distinguish properly between uncontrollability and unpredictability (Kudielka, Hellhammer, & Kirschbaum, 2007; Smeets et al., 2012) probably because they are often related.

***Emotional stressors*.** Negative feelings in an emergency situation are, for example, fear, anxiety, a feeling of helplessness, and frustration (Nass et al., 2015). Emotions could be seen as accompaniment of emergency situations (Nass et al., 2015), but also as a stressor (Dickerson & Kemeny, 2004). In our study, emotions were used as a stressor but also seen as accompaniment of the simulated emergency situation. Emotions could be evoked by applying stimuli from databases, which are mentioned under references in table 1. The stimuli in the databases were rated on the Self-Assessment manikin (SAM) scale (e.g. Bradley & Lang, 1999; Lang et al., 1997). The SAM ratings of pictures are very similar between US citizens and Brazilians (Ribeiro, Pompéia & Amodeo Bueno, 2005). This is an indication that other cultures rate the pictures similarly. The ratings of pictures revealed that some pictures evoke only one emotion while other pictures evoke multiple emotions (Mikels et al., 2005).

Cognitive stressors. Cognitive tasks, such as mental arithmetic tasks, function as stressors, even if the participants' performance on the task was not evaluated by others (Dickerson & Kemeny, 2004).

Cognitive tasks are often applied in laboratory stress tests (e.g. TSST: Kirschbaum et al., 1993; MMST: Kolotylova et al., 2008). Some authors (Brouwer & Hogervorst, 2014; Niculescu, Cao, & Nijholt, 2008) assume that the cognitive tasks evoke stress through increasing workload. The transactional stress model emphasis that stress is evoked when the demands exceed the capabilities. Therefore it is reasonable, that stress is evoked if cognitive tasks increase cognitive load up to a certain level.

Acoustic stressors. Talking at a normal volume is around 60 dB(A), but even noise levels of 55 decibels are enough to make communication more difficult (Bahr et al., 2015). When people are exposed to 80 decibels constantly, the noise can impact health. Above 130 decibels even a short exposure time is enough to harm the sense of hearing. In laboratory studies, acoustic stressors of 90 db(A) (Miki, Kawamorita, Araga, Musha, & Sudo, 1998) and increasing from 78 to 93 db(A) (Kolotylova et al., 2008) were applied. The study by Miki et al. (1998) found an interaction effect of an acoustic and a cognitive stressor. Completing an arithmetic task (cognitive stressor) only had a small effect on cortisol under quiet conditions but in a noisy condition much more cortisol was released.

Several conclusions can be drawn from the presented literature. Firstly, there are different categories of stressors. According to Dickerson and Kemeny (2004), the categories differ in their effectiveness of evoking stress. Secondly, also the stressors within a category differ in their effectiveness. Thirdly, the meta-analysis by Dickerson and Kemeny (2004) and the study by Miki et al. (1998) about acoustic stressors indicate that combinations of stressor categories are beneficial. Many laboratory stress protocols, such as the well-known Trier Social Stress Test (TSST; Kirschbaum, Pirke, & Hellhammer, 1993) and the Mannheim Multicomponent stress test (MMST; Kolotylova et al., 2008), already combine several stressor categories. For instance, the MMST combines among others an emotional stressor, cognitive stressors and acoustic stressors.

1.2.2. Stress measures

Stress can be measured in several ways. This paragraph discusses objective and subjective measures and users' performance metrics.

Objective measures. There are several possibilities to measure stress objectively. The most common physiological measures are heart rate measures (e.g., Hjortskov et al., 2004; Reinhardt et al., 2012), endocrine measures, and skin conductance (e.g., Brouwer & Hogervorst, 2014; Fernández et al., 2012; Niculescu et al., 2010; Reinhardt et al., 2012). The heart rate measures and the skin conductance measures are not appropriate for this study since it is hard to tell stress apart from physical activity, such as running. Studies using heart rate measures analysed the heart rate variability (HRV), calculated the ratio of low frequency and high frequency (LF/HF ratio) (Hjortskov et al., 2004), calculated the mean heart rate (beats per minute) or mean peak heart rate (Reinhardt, Schmahl, Wüst, & Bohus, 2012).

Endocrine stress measures analyse the alpha-amylase level (Takai et al., 2004) and cortisol level (e.g. Bigert, Bluhm, & Theorell, 2005). The focus is put on cortisol measures, because it is the most often used and most discussed endocrine measure. Cortisol samples can be obtained from saliva, urine, or blood serum (Bigert et al., 2005). Salivary cortisol has the advantage that it is not necessary to store it at a certain temperature until analysis (Bigert et al., 2005). A disadvantage of measuring cortisol is that some factors influence the amount of cortisol (Hellhammer, Wüst & Kudielka, 2009). Those factors are among others the time of day (Hellhammer et al., 2009), medical conditions, smoking, caffeine and alcohol use, exercise and recent meals (Adam & Kumari, 2009). Cortisol decreases during the day (Raff & Trivedi, 2013). Furthermore, several studies found variations in cortisol levels between healthy individuals and individuals with psychological disorders (e.g. Yehuda, Teicher, Trestman, Levengood, & Siever, 1996). Physiological diseases, such as a renal disease, also cause variation in cortisol (e.g. Raff & Trivedi, 2013). Even genes could moderate the amount of cortisol induced by some stressors (Allen, Kennedy, Cryan, Dinan, & Clarke, 2014).

Cortisol is released when the activity of the sympathetic nervous system declines (see description in 1.2). Therefore, the peak in the cortisol response to a stressor is delayed. The presented stress protocols vary in delay. The MMST reports a delay of twenty minutes (Reinhardt et al., 2012) and the TSST a delay of forty-five minutes (Kirschbaum et al., 1993) after the start of the experiment. According to Dickerson and Kemeny (2004) not all stressor categories are able to evoke the release of cortisol. Their meta-analysis did not show an effect of emotional and acoustic stressors. However, other studies did find an effect (e.g. Fernández et al., 2012; Miki et al., 1998). The MMST and TSST as well as other stress tests combined several stressor categories. Therefore, we assume that a combination of stressors evoke the release of cortisol.

Subjective measures. Subjective stress measures were in line with the objective measures in several studies (e.g. McRae et al., 2006; Reinhardt et al., 2012). Subjective stress ratings could be obtained by marking the stress level on a visual analogue scale ranging from zero to one-hundred (Hellhammer & Schubert, 2012) or indicating the stress level by means of a Likert scale ranging from zero to nine (Reinhardt et al., 2012) or to ten (Brouwer & Hogervorst, 2014). Both types of measurements, the line and the Likert scale, lead to similar results (Carifio & Perla, 2007). How long the stress experience lasts depends on the stress manipulation. Five to thirty minutes are reported by McRae et al. (2006).

The Self-Assessment Manikin (SAM) could be applied to measure emotions (e.g. Bradley & Lang, 1999; Lang et al., 1997). It consists of three scales: the valence scale assessing how happy people feel, the arousal scale assessing how excited participants feel, and the dominance scale assessing how much they feel under control of the situation.

User's performance metrics. The presence of stress can be indicated by analysing the participants' voice. A study by Niculescu et al. (2010) used verbal hesitation, breaks, mispronunciation, and number of words as measures to indicate stress. There is even an application to report the level of stress through an analysis of the voice (Lu et al., 2012). Stress could also be indicated by performance

decrements. Oei et al. (2006) stated that performance in a reaction task and memory recall task decreased when cortisol was high. However, neither the app nor the other speech analyses were applied because a proper voice recording during the experiment cannot be assured, since voice recording apps may interfere with the RESCUER App. The memory recall task could be a possibility to measure stress.

1.3. Moderator Variables

The variables moderating the cortisol response were already presented in 1.2.2. The variables moderating the performance and stress perception in an emergency situation are presented here.

Some papers discuss the relation of personality and stress (Mogg, Bradley, & Hallowell, 1994; Stankovic, Fairchild, Aitken, & Clark, 2014; Vollrath, 2001). The study by Mogg et al. (1994) found a relation of stress and trait anxiety, which affected the direction of attention to threat words. Stankovic et al. (2014) found a relation of stress and trait impulsivity in a risky choice task. Personality also influences the coping strategies (Penley & Tomaka, 2002). Some people benefit from medium or high emotional intelligence when coping with stress (Gohm, Corser, & Dalsky, 2005). To sum it up, personality could account for individual differences in the stress response.

The participants' age may affect the performance in a stress experiment since some brain regions involved in information processing mature late and are affected by aging. For instance, the prefrontal cortex (PFC) is responsible for decision making and working memory (Glendon, 2011). It is one of the last brain regions to mature (Glendon, 2011), and its functionality declines in almost all older adults (Schultheis & Manning, 2011). Furthermore, walking affects dual task performance in older people more than in younger ones (Srygley, Mirelman, Herman, Giladi, & Hausdorff, 2009).

Previous experience with emergencies may affect the experiment. Experiences are stored in long-term memory, which could influence other parts of information processing (Wickens, Lee, Liu, & Gordon Becker, 2004). In one evaluation of the RESCUER App, an interaction effect was found between ratings of cognitive load and prior experience with emergency situations (Nass, Jung, Groen, & Villela, 2015b). The participants with experience rated the app screens as more difficult in comparison to participants without experience.

Some moderator variables could be excluded by the selection of the participants. Selection criteria are age, no physiological impairments that affect using an app and walking, and no severe psychological disorders. Other variables have to be assessed by a questionnaire, such as experience with emergencies, or be assessed by other means.

1.4. Creating the Simulation

The simulation consists of the stress manipulation and the simulation of the use context. The basis for the simulation of the use context are the requirements (see 1.1.) . This part of the simulation has to be adapted to the system, the available location, budget and participants. The stressors only need

adjustments to other crowd-based rescue systems or another evaluation if the incident type is different. For instance, when the incident is a fire the sound of a smoke detector can be used as an acoustic stressor. In case of an earthquake other sounds make more sense. The simulation of the use context for the RESCUER App and the stress manipulation are described in the following, as well as some practical issues that should be considered.

1.4.1. Creating the use context for the RESCUER App

The use context (see 1.1.) and table 1 described the requirements for simulations for crowd-based rescue systems in general. In the following section, a simulation is described that is adapted to the evaluation of the RESCUER App.

In the experiment, participants have the role of eyewitnesses who experience an incident up close. We chose to focus on eyewitnesses, since there are normally more eyewitnesses than first responders present in an emergency situation. The number of participants was limited to fifteen participants at the same time, due to security reasons and available materials. Only some participants were provided with a smartphone on which the RESCUER App was installed. Since it was not possible to do the experiment in an industrial park or at a location of a large scale event, the experiment took place at Fraunhofer IESE. The whole experiment took place inside the building because the risk of black ice at the time of conducting the experiment made outside areas unsafe. The place of the simulated incident and the safe area were as far apart from each other as possible. The start point was a meeting room and the final place of the experiment, referred to as Point B, was one too. Meeting rooms were chosen so that the participants could fill in questionnaires and listen to instructions while sitting. Going to Point B, the participants passed through the three zones, which were simulated by the number of stressful stimuli. In the hot zone, most stressors were presented, and in the cold zone, no stressors at all. The distance between the rooms could be covered in around three minutes when walking slowly. The participants were told to hurry but they were not told to run (the reason for this is explained in 1.5.). For evaluating the app it was important that they walk since walking could impact the performance on cognitive tasks (Srygley et al., 2009). At least one area on the way to Point B was unknown to all participants, since this part had been under construction and was not reopened yet. At best, attending a large scale event or being at an industrial park should be simulated. However, the number of people and the location did not fit those settings. For this reason, a scenario was created that was more suitable, namely being eye-witnesses of an explosion while attending a seminar. An explosion was chosen since this incident type had not been evaluated yet. The scenario was created by text, a video, and pictures on posters. The video and pictures were the same in the control and experimental group so that the reported information could be compared. Therefore, the video and pictures did not contain any negative emotional stimuli, such as injured persons or faces expressing fear. There were other posters which presented emotional stimuli and neutral stimuli because solely presenting relevant information on posters would make relevant stimuli more obvious than in a real emergency situation, where individuals also have to extract relevant stimuli from many others.

1.4.2. Stressors for the simulation

From the section about stress (see 1.2.) we learnt that a combination of stressor categories should be applied that make use of visual and acoustic channels. Olfactory and thermal stressors could also be considered but they are not applicable in the given location. The selected stressors are presented in figure 2 and are explained in more detail in the method section (see 2.4.).

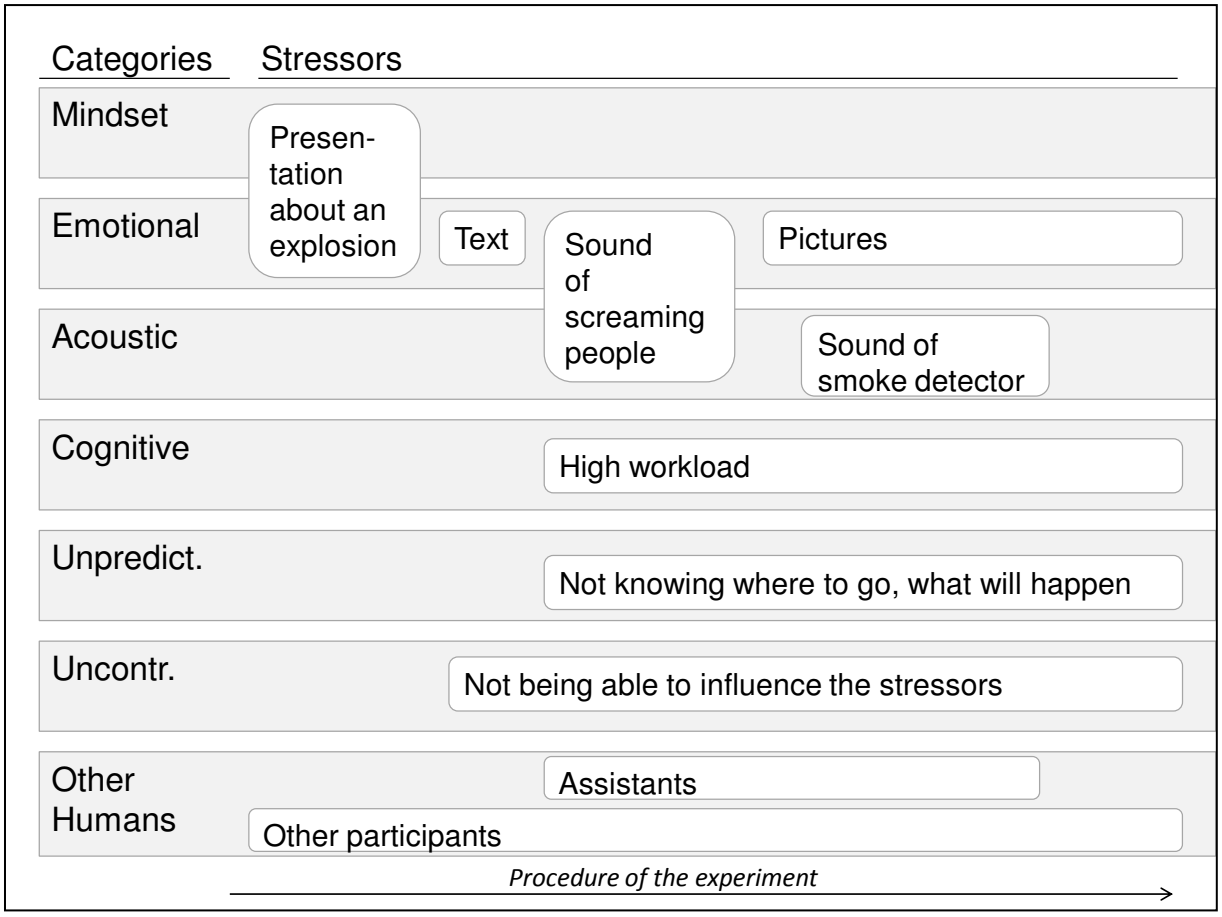


Figure 2. Overview over the applied stressor categories (grey boxes) and the related stressors (white boxes). Some stressors could be assigned to two categories. The order of the stressors represents the order during the experiment. Several stressors were present at the same time. In those cases, interaction effects are possible.

1.4.3. Issues to take into consideration

The set-up of the simulation has to take into account that stress and emotion affect perception (Mendl, 1999; Schwabe & Wolf, 2010). Stress may cause attentional tunnelling, that means the range of attention is restricted (Wickens et al., 2004). Peripheral cues are not perceived so that the focus is on the more important cues (Ozel, 2001). Thus, participants may miss visual stimuli because their focus is more on cues helping them to find the way to safety. To counteract this, the visual stimuli should be positioned on prominent places and the participants should be instructed to actively look for them. Moreover, people may forget that they have the rescue system with them (Bergstrand & Landgren, 2009). In the simulation, the participants should be reminded about using it.

To analyse whether the participants really experience stress, a control group was necessary that experiences stress not at all or only mildly. The procedure should stay the same, but all negative stimuli should be replaced by positive ones, for instance pictures of happy persons and relaxing music instead of sad persons and the sound of a smoke detector. A presentation at the beginning of the simulation is supposed to bring the participants in a mindset that facilitates emphasizing with the scenario. In the control group, the presentation functions as a protective frame (Fokkinga & Desmet, 2013), which enables the participants to enjoy an explosion. In the simulation applied to the RESCUER App the frame was created by putting the explosion in a controlled setting in which it is assumed to lose its harmfulness.

1.4.4. Pilot study

In a pilot study, two people represented the control group and two the experimental group. One person in each group was provided with the app. In the experimental group, the experiment started with a presentation about explosions to bring the participants into the right mindset, and then they read a text saying that they are close to an explosion. The text contained eleven bold printed words which were related to emergencies or negative emotions. Next they watched a video of an explosion and then went back to the text to memorize the bold printed words within two minutes. The memorizing task was meant to increase the cognitive workload and to keep emotional words in mind during the experiment. After two minutes, they went to Point B while listening to the sound of fire alarm, which was played from the smartphone for practical reasons, and reporting the incident via the app. Along the way, there were photos they had to report. The procedure of the control group was the same but with positive stimuli, such as relaxing music instead of the sound of fire alarm.

The pilot study revealed that the participants forgot the words too fast to keep cognitive load high during the experiment. Therefore, the memorizing task was removed. The new approach was then to include more stimuli that are present in an emergency situation to make the simulated emergency situation itself cognitively demanding enough. New stimuli were assistants who stood along the path to Point B and shouted at the participants to hurry and acted hectic. Other added stimuli were one confederate in each group who behaved in a stressed manner in the experimental group and calmly in the control group. Those stimuli were based on the assumption that people adopt emotions of others (Nass et al., 2015). Moreover, the pilot study revealed that the stressors have to be closer in time to keep the stress level constant. Another pilot study with an improved stress manipulation was successful in evoking constant stress.

1.5. Research Question and Hypotheses

This thesis had the purpose of answering the following research question:

Is the presented simulation of an emergency situation useful for evaluating crowd-based rescue systems?

Two subquestions had to be answered. Namely, was the stress induction successful and was the simulation of the use context successful?

The success of the stress manipulation was reflected in the data in three ways; cortisol level, subjective stress ratings and ratings on the Self-Assessment Manikin (SAM). The cortisol level was assumed to increase from the first to the second measure in the experimental group. In the control group the difference between the two measures was assumed to decrease slightly according to the normal day schedule. The subjective stress level was hypothesised to be higher during the experiment in comparison to before the experiment in the experimental group. The control group was assumed to only indicate mild levels of stress during the experiment and only a small difference compared to before and after the rating. The participants in the experimental group were expected to feel less happy, less under control, and more aroused during the experiment than before the experiment (Dickerson & Kemeny, 2004; Nass et al., 2015). Their absolute ratings were assumed to reflect unhappiness, low control and high arousal during the experiment. The control group was expected to feel as happy and under control of the situation during the experiment as before and after the experiment. The control group was expected to feel aroused during the experiment, too, since those participants had to be active during the experiment. However, the increase in arousal was expected to be lower than in the experimental group. The control group was assumed to feel medium to high levels of valence and dominance. Furthermore, we assumed that the mean ratings of the stressors reflect at least medium stressfulness.

The simulation was considered to be successful, when the participants empathized with the scenario. Empathizing took place when they walked to the safe area while reporting the incident, they perceived the stressors, and they were not distracted by anything not related to the experiment. It was expected that the participants in the experimental group would run or at least walk fast if they feel that it is necessary to flee from the incident. The assistants did not tell them to run so that it could be observed how the participants behaved because of the simulation. The control group on the other hand was expected to walk at a normal pace.

2. Method

2.1. Participants

Twenty-two persons participated in this experiment, ten of them in the control group. The participants' age ranged from 22 to 31 (M = 26.18, SD = 2.13). In the experimental group 83.3 % were male, in the control group 50%. All participants were students, 81.1% of them had already achieved a bachelor or master degree. In the control group, five students were related to Fraunhofer IESE by writing their thesis at the institute and/ or working there as a research assistant. One of those students in the control group already knew what the RESCUER project was about before the experiment. In the experimental group, four persons were related to Fraunhofer, of whom one was involved in the RESCUER project. The others were students at TU Kaiserslautern, and received a credit point for participation. The

participants' study program was in English. The participants were recruited by sending e-mails and by personally asking students related to Fraunhofer IESE. The invitation informed about the exclusion criteria (physiological or psychological impairments) and from which activities they should refrain from. All participants indicated to have refrained from consuming food, beverages containing alcohol or caffeine and extensive exercise at least one hour before the experiment and half an hour before from any beverages. Personality, emotional awareness and genetics were not controlled since obtaining them would consume too much time and budget. Several nationalities were represented in the sample. In the control group, 50% were Iranian, 30% German, and 20% Indian. In the experimental group 50% were Indian, 25% Brazilian, 16.7% Pakistani, and 8.3% Iranian.

None of the participants indicated to be physically or mentally handicapped in a way that could have influenced the experiment. Six of the participants had experience with emergencies. An emergency was defined as a situation that required calling rescue workers.

2.2. Materials

Different types of smartphones with the Android operating system were used (Archos 50 Platinum, 2x LG Nexus 2, LG Nexus 5, Samsung Galaxy S3, and Samsung Note 2). The latest version of the RESCUER App was installed, called rescuer_0.42.0.r425. This version was used without WiFi because of the network instability in certain parts of the escape route. The data was saved on the phone and was sent via e-mail to a computer for analysis after the experiment. The app icon was on the start screen. The video showing an explosion used was derived from YouTube (YouTube, 2015). The presentations and the texts were accompanied by music. In the experimental group a sad and dramatic piece of music was used (Secession Studios, 2014) and in the control group happy music (YouTube, 2013) was used. The statistical analyses were conducted using IBM SPSS 23.0.

2.3. Scenario and Task

The scenario was that the participants were attending a seminar (described by a text) when an explosion happened on the street (video). They had to flee to the safe place Point B. On the way, they saw destroyed houses and fleeing persons (posters). There were also neutral posters, depicting houses, trees and cars. Within this scenario, the participants had to complete two tasks. The first task was to go from the meeting room to point B. The second task was to keep their eyes open for information about the incident presented along the way to point B and reporting the information. The app-users were instructed to start with the report on the way to point B and the others to report the information by filling out a questionnaire at point B.

2.4. Stress Manipulation

Stressors belonging to the categories (mentioned in 1.4.2.) were applied. Figure 2 provides an overview over the stressors and their chronological order. The experiment was approved by the ethics committee at the University of Twente.

The presentation (appendix F) at the beginning of the simulation was about the fireworks disaster in Enschede in 2000. The presentations did not contain any signs of fire so that the participants could not confuse the disaster with a fire. The emotional stressors were a text and pictures. The text (appendix H) described the scenario and included bold printed words, which were related to emotions or emergencies (e.g. sad, harmful). It was planned to use words of the Affective Norms of English Words (ANEW: Bradley & Lang, 1999), however, the institute responsible for providing the ANEW did not respond to requests to make the words available for research purposes. The pictures (appendix K) showed a crying child and people running away from a threat and were in approximately A2 format attached to the walls along the path to point B. Acoustic stressors were the sound of a fire alarm played at the volume of 80dB and the sound of screaming people. The sound of screaming people was also an emotional stressor.

A cognitive stressor was not included for the reasons described in '1.4.4. pilot study'. The situation itself was considered to be cognitively demanding, because the participants had to complete several tasks at the same time; they had to keep the information from the video active in their memory, they had to look for information in their surroundings, extract relevant information and report it through the app or keep it in working memory, and they had to listen to the instructions of the assistants and find their way through the building. Uncontrollability was part of the experiment since the participants could not influence the unpleasant stimuli such as the sound of the smoke detector. Unpredictability was evoked by not providing information about the duration of the experiment, the exact way to point B, and what was going to happen during the experiment (see procedure). A feeling of uncontrollability and unpredictability was also described in the text about the scenario. Furthermore, when the participants left the meeting room, they entered a dimly lit corridor which was assumed to make it more difficult to get an overview over the situation.

The behaviour of the assistants along the way to Point B and the confederate among the participants was meant to evoke stress and negative emotions in the participants by behaving erratically and not smiling. The assistants informed the participants about the direction to point B by shouting "Go in this direction" and they told them to hurry (see appendix J for their instructions). Telling them the direction assured that the participants walked along the pre-defined way. Still, the participants had to orient themselves to find the way to point B. Moreover, participants expressing emotions may act as emotional stimuli, too (Nass et al., 2015).

In the control group, the presentation was about the Mythbusters, an American TV program combining science and entertainment who had fun blowing up a car in a controlled setting. The control group used the same categories of stimuli but in a no/low stress version. The text said that a

Mythbuster experiment went slightly wrong but the participants were at a safe distance, neutral or positive words were highlighted (appendix I). The presentation and the text describing the scenario were accompanied by happy music (YouTube, 2013). After the text the sound of laughing people was played at the volume similar to a conversation. On the way to point B, relaxing music was played (Explosions in the Sky, 2004). Instead of posters depicting negative stimuli, happy persons were depicted. The assistants along the way were smiling and informing the participants where to go in a calm and friendly way.

2.5. Procedure

The procedure is described from the perspective of the experimental group. The control group followed the same procedure but with the non-stressful stimuli. Figure 3 illustrates the procedure. The control group started at 10 a.m. and the experimental group in the afternoon at 3 p.m. Each experiment lasted about 80 minutes. In each group, there was supposed to be one confederate and twelve participants. The number of participants was calculated with a g*-analysis (appendix B). However, in the control group, two persons did not show up resulting in ten participants. Appendix C presents the protocol of the experiment with more details about the procedure and an explanation of the justifications of the procedure.

A single experiment leader guided the participants through the procedure. The participants were welcomed and asked to take a seat. All participants randomly drew a piece of paper, which decided whether they were provided with a smartphone with the RESCUER App or not. Then the participants signed the informed consent (appendix D), which was in line with the ethical requirement for stress research (appendix E). Thereafter, a short presentation was given about psychological experiments in general with the purpose of reducing social desirability and potential testing effects. Then the first saliva sample was taken. After that, the participants filled in a questionnaire (appendix G). The questionnaire asked demographic questions, the self-reported current stress and emotional level, and information about potential moderator variables, such as experience with emergencies. After completing the questionnaire, the RESCUER App was introduced and the tasks the participants had to complete were explained. Then, the participants watched the presentation about an explosion. Thereafter, a short reminder of the scenario was provided that was directly followed by the text and the video. When the video ended, the sound of screaming people was played. The participants left the room to go to point B. On the way, they performed the task. When all participants had arrived at point B, they were told that they could relax and finish reporting through the chat function. The duration from the beginning of the presentation to arriving at point B took about five to six minutes. Then, the participants were asked to sit down and fill out the second questionnaire (appendix G). Twenty-five minutes after the stress manipulation started, the second cortisol sample was taken. The participants who did not finish the questionnaire by this time could continue after taking the sample. The smartphones were collected while the participants filled out the questionnaire. The tape with the

participant number was stuck to the phones to later identify the user of the phone. When all questionnaires were filled out, the participants in the experimental group were debriefed directly and the control group was informed that they would receive an e-mail the next day. They were not debriefed directly so that they could not pass on any details about the experiment to the participants who participated later that day as part of the experimental group. All participants were thanked for participating. The control group was served coffee, which was announced days before the experiment so that they had an incentive to refrain from caffeine beverages in the morning.

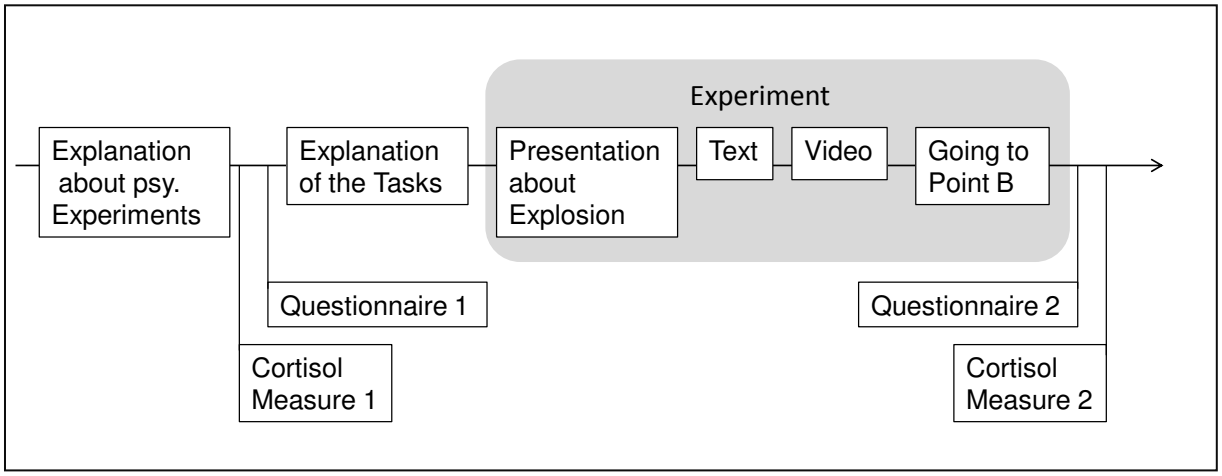


Figure 3. Overview of the most important steps of the experiment. Inside the grey box, the steps of the experiment are depicted.

2.6. Measures

Subjective measures. The participants indicated their stress level on a Likert scale ranging from zero to ten. This type of scale was chosen because it takes less effort to analyse than a line ranging from zero to one hundred. The Self-Assessment Manikin (SAM) was not used to determine the exact emotion but to analyse whether the participants felt accordingly to an emergency situation. SAM was introduced in questionnaire 1 (appendix G) with a short explanation about the scales. All three scales (valence, arousal, and dominance) were applied. A version of SAM was used that depicted circles under the nine pictures to facilitate marking the SAM pictures. Only in the questionnaire for the experimental group without app, the circles were mistakenly not depicted. The subjective measures were obtained for three moments in time: before, during, and after the experiment. The measures for before and during the experiment were obtained at the same time, namely when filling in the second questionnaire. Asking the participants directly during the experiment to rate stress and SAM would have interrupted the flow of the experiment. We assumed that a few minutes after the experiment ended, they were still able to express how they felt during the experiment.

Objective measures. Stress was measured objectively by analysing salivary cortisol samples. Salivary cortisol was chosen because it does not require the assistance of medical staff as blood cortisol does and it is easier to obtain at a certain point in time than urinary cortisol. Saliva cortisol was obtained by using the Medivera® saliva kit. The samples were taken twice; about 20 minutes before the

experiment started and 25 minutes after the experiment ended. The measure before the experiment was obtained before the participants faced any stimuli related to emergencies. The time of the peak cortisol level differs between studies (e.g. Kirschbaum et al., 1993; Kolotylova et al., 2008). The time of the second measure was inspired by the peak level in the Manheim Multicomponent Stress Test (MMST), because the duration of the stress manipulation in the MMST (5 min) is similar to our study (5-7 min) and choice of stressors is similar to MMST, which also combined an acoustic stressor with psychological stressors. The second cortisol measure was supposed to be 20 minutes after the experiment, but the actual time was 25 minutes so that more participants could finish the second questionnaire before the measure. The cortisol analysis was conducted by Medivera®.

Other measures. The assistants and the experiment leader reported their observations. They did not follow an observation protocol since this would have distracted them from their main task. Additionally, the perceived stressfulness ratings of the stressors and other elements of the experiment that may have unintentionally evoked stress were obtained from the participants. The SAM dominance ratings gave an indication for the success of the stressor uncontrollability.

2.7. Data analysis

A linear mixed model was applied since it is a convenient way to deal with repeated measures. The cortisol was measured two times, and the subjective stress ratings and the SAM ratings three times. The predictors in all analyses were the group, the app use, and the moment of measure. Interaction effects between all variables were included.

From the stress elements, only the means were calculated to get an impression of the perceived stressfulness of the elements. Further analyses, such as comparing the means, was not conducted since the design of the experiment did not allow for the exclusion of interaction effects. The success of the simulation was determined based on the reported information through the app, the observation of the participants, and the reports of the participants and a confederate about how they perceived the experiment in an open interview.

3. Results

The tables about the descriptive statistics and the variance are presented in appendix M, the testing of assumptions in appendix N, and the syntax in appendix O.

3.1. Cortisol measures

Figure *4* compares the two measured cortisol levels with regard to the group. The diagram shows that one or both cortisol measures of five participants were extremely low; below 500 pg/ml. These levels could be due to a measurement error or were an indication of severe health issues. For this reason, those participants were excluded from the statistical analysis. The diagram provides information on whether cortisol increases or decreases from the measurement before the experiment (baseline) to after

the experiment. An increase implied an increase in stress; a decrease implied that the participants experienced no stress or their stress level decreased. The diagram shows that in both groups the cortisol level of most participants decreased from before to after the experiment. Only five participants showed an increase in cortisol, two in the control group and three in the experimental group. That means the cortisol measures did not support the success of the stress manipulation. The statistical data supported this finding (table 3). The interaction effect of moment and group is small while at the same time the confidence interval is large (B = 206.03, 95% CI = [-2901.56, 3313.62]).

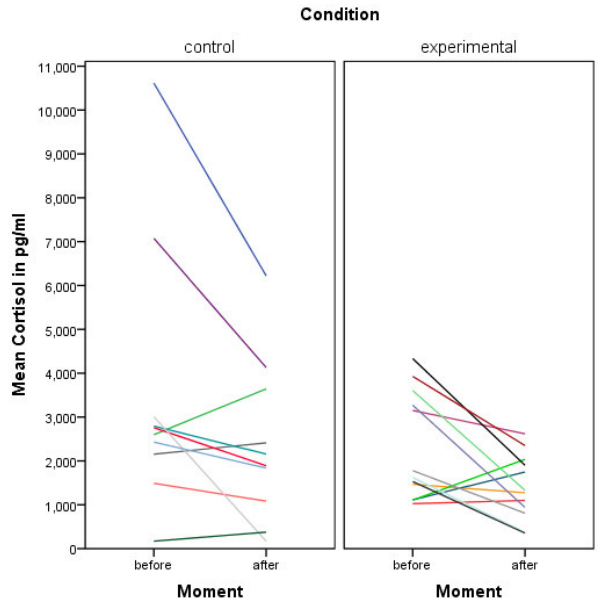


Figure 4. Depiction of the individual results of the cortisol measures . The moment of measure is with regard to the experiment. The results are distinguished by the group the participants were in. The control group is depicted on the left and the experimental group on the right.

Table 3

Estimates of Fixed Effects of Cortisol Measures Excluding Participants with Extreme Values

Parameter			95% Confidence Interval	
	Estimate	Sig.	Lower Bound	Upper Bound
Intercept	1632.00	<.01	711.89	2552.11
[Moment=before]	1176.17	.26	- 918.97	3271.30
[Group=Control]	2026.40	.01	661.65	3391.15
[App = without]	-55.00	.94	-1509.82	1399.82
[Mom=before] * [Group=Control]	206.03	.89	-2901.56	3313.62
[Mom=before] * [App= without]	-770.42	.63	-4083.12	2542.29
[Group=Control] * [A = without]	-1911.07	.08	-4107.80	285.67
[Mom=before] * [Group=Control] * [App=without]	-67.12	.98	-5069.19	4934.96

Note. N = 17

3.2. Subjective Ratings

Figure 5 presents the stress ratings. The value zero means not stressed at all and ten means extremely stressed. Figure 5 shows that all participants indicated an increase in stress from before to during the experiment, except for two participants. The mean increase for the control group with the app (M = 4.60, SD = 1.52) is as high as in the experimental group without the app (M = 4.60, SD = 3.13).

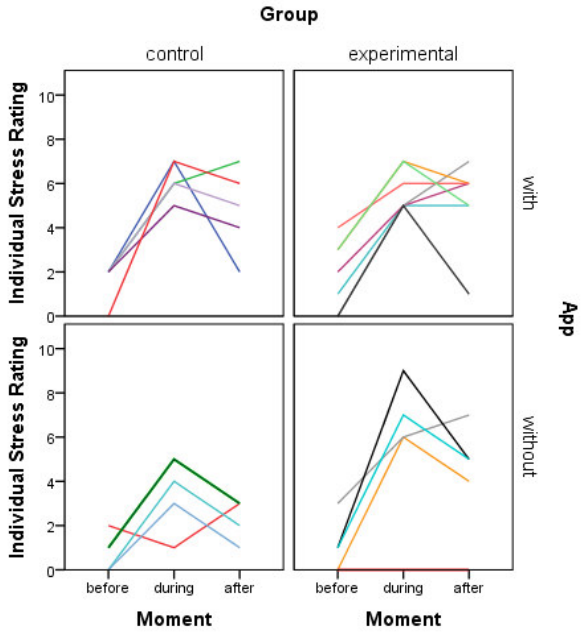


Figure 5. Individual stress ratings sorted by group and app use.

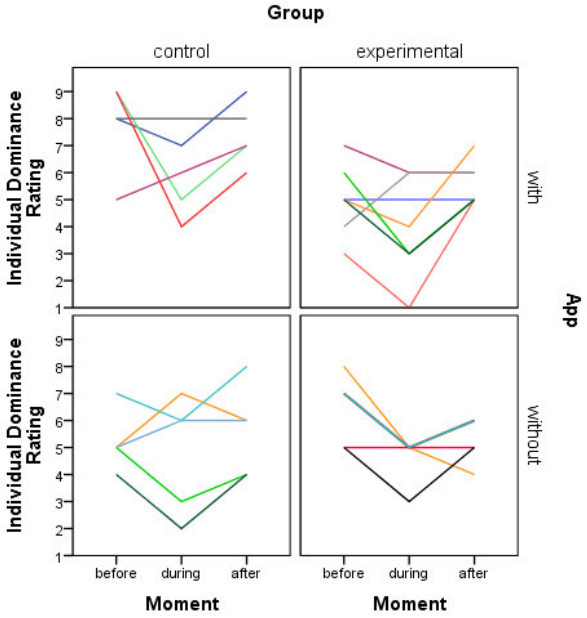


Figure 6. Individual dominance ratings sorted by group and app use.

The statistics confirm the effect of the moment (B = -3.00, 95% CI = [-5.24, -0.76]). The participants in the control group without the app indicated only low to moderate stress (range = [1, 5]) and the other participants moderate to high levels (range = [5, 9]). Table 4 confirms that the control group indicated lower stress ratings and that there is an effect of the 2x2 groups (= app*group) but shows only a small effect of the interaction of the group and the moment. The assumption that the experimental group shows an obvious increase in stress during the experiment and the control group only a mild increase compared to before and after the experiment is not confirmed since it was not expected that the control group indicated such an increase in stress and that the control group without the app showed similarly high ratings as the experimental group.

Figure 6 presents the dominance ratings. It shows that there was variation in the dominance ratings. Four participants indicated that they felt the same level of control at all three moments of measure, four indicated that they felt more in control during the experiment than before, and the rest indicated that they felt less in control during the experiment compared to before. The moment is a predictor of the ratings (B = 1.20, 95% CI = [-0.04, 2.44]). It predicts that the measure during the experiment is lower compared to before and after the experiment.

Table 4

Estimates of Fixed Effects of the Subjective Stress Ratings

			95% Confidence Interval	
Parameter	Estimate	Sig.	Lower Bound	Upper Bound
Intercept	5.60	<.01	3.72	7.48
[Moment_bin=.00]	-3.00	.01	-5.24	-0.76
[App=with]	0.11	.93	-2.35	2.58
[Group=Control]	-2.00	.14	-4.66	0.66
[Moment_bin=.00] * [App=with]	0.79	.59	-2.15	3.72
[Moment_bin=.00] * [Group=Control]	1.00	.53	-2.17	4.17
[App=with] * [Group=Control]	2.49	.18	-1.14	6.11
[Moment_bin=.00] * [App=with] * [Group=Control]	-1.79	.41	-6.10	2.53

Note. Moment_bin=.00 is a combination of the measure before and after the experiment.

Table 5

Estimates of Fixed Effects of the Subjective Dominance Ratings

			95% Confidence Interval	
Parameter	Estimate	Sig.	Lower Bound	Upper Bound
Intercept	4.60	<.01	3.32	5.88
[Moment_bin=.00]	1.20	.06	-0.04	2.44
[App=with]	-0.60	.47	-2.27	1.07
[Group=Control]	0.20	.82	-1.61	2.01
[Moment_bin=.00] * [App=with]	0.09	.92	-1.54	1.71
[Moment_bin=.00] * [Group=Control]	-0.60	.49	-2.36	1.16
[App=with] * [Group=Control]	1.80	.15	-0.66	4.26
[Moment_bin=.00] * [App=with] * [Group=Control]	0.91	.44	-1.48	3.31

Note. Moment_bin=.00 is a combination of the measure before and after the experiment.

In the control group with app, the participants indicated that they felt strongly in control of the situation before the experiment. In the other groups there is more variability in the data at this moment. Several participants in the experimental group rated dominance during the experiment with 5 or higher. Lower ratings were expected. There is only a very small effect of the group (B = 0.20, 95% CI = [-1.61, 2.01], table 5) and the interaction of the group and moment (B = -0.60, 95% CI = [2.36,

1.16]). However, the group in combination with the app use had an effect on the ratings (B = 1.80, 95% CI = [-0.66, 4.26]). The control group with the app felt like they had more control, than the other group and app use combinations. The hypothesis that only the experimental group felt less under control and low in the absolute control rating during the experiment is not confirmed.

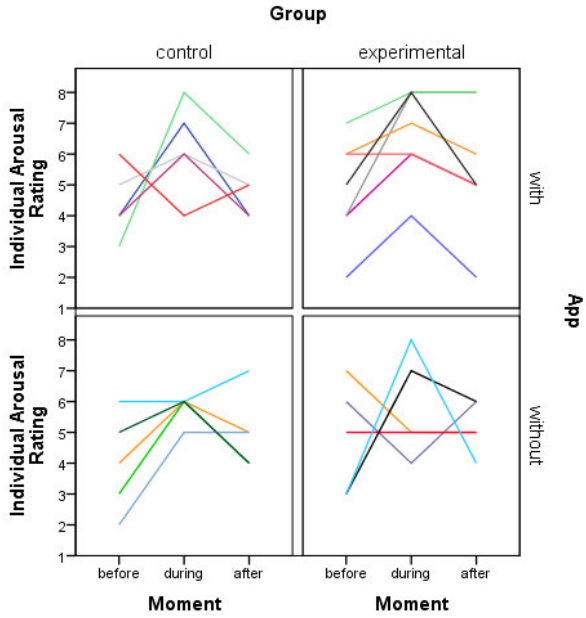


Figure 7. Individual arousal ratings sorted by group and app use.

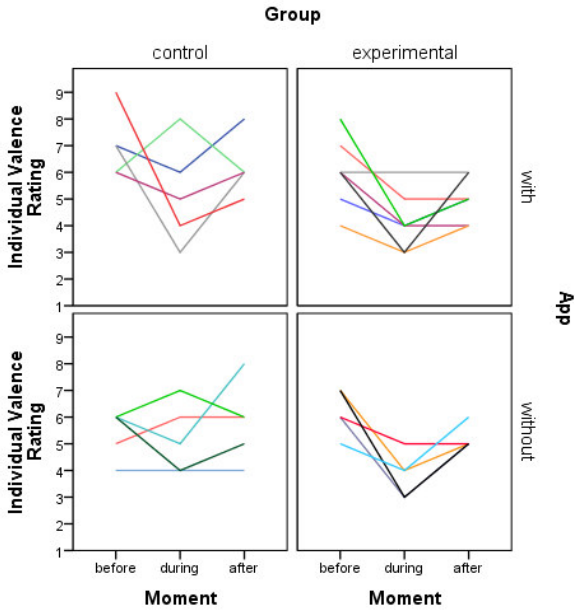


Figure 8. Individual valence ratings sorted by group and app use.

Figure 7 presents the individual arousal ratings. All participants except for five indicated to feel more aroused during the experiment compared to before and after. However, table 6 shows only a small effect of the moment of measure on the ratings as well as of the interaction of moment and group. The ratings in the control group without the app do not exceed six during the experiment. In the other app and group combinations the maximum arousal ratings were eight, which represented a high level of arousal. All predictor variables only have a small effect on the ratings (see table 6). Figure 7 shows that the assumption that both groups experience more arousal during the experiment was true for most participants. However, the assumption that the control group experienced less arousal than the experimental group is not confirmed (B < 0.01, 95% CI = [-1.85, 1.85]).

Figure 8 presents the valence ratings. In the experimental group, all participants except for one indicated a clear decrease in happiness during the experiment in comparison to before and after the experiment. The statistics state that the ratings during the experiment were lower (B = 1.90, 95% CI = [0.67, 3.13]). The control group had higher ratings than the experimental group (B = 1.40, 95% CI = [-0.09, 2.89]). The control group with the app had the highest happiness ratings before the experiment. Table 7 shows that there was a similar strong interaction effect of moment, app use and group and of moment and group. The control group had a smaller difference between the ratings before and after the experiment compared to during. The hypothesis stated that only the experimental group felt not happy

during the experiment. However, the data shows that there were participants in the control group who felt unhappy and less happy compared to before and after the experiment. On the other hand, the experimental group indicated a larger decrease in happiness during the experiment and the control group had in general higher ratings. Therefore, the hypothesis is only partly confirmed.

Table 6
Estimates of Fixed Effects of the Subjective Arousal Ratings

				95% Confidence Interval	
Parameter	Estimate	Std. Error	Sig.	Lower Bound	Upper Bound
Intercept	5.80	0.65	<.01	4.49	7.11
[Moment_bin=.00]	-0.80	0.69	.25	-2.20	0.60
[App=with]	0.91	0.85	.29	-0.80	2.63
[Group=Control]	<0.01	0.92	1.00	-1.85	1.85
[Moment_bin=.00] * [App=with]	-0.70	0.91	.44	-2.53	1.13
[Moment_bin=.00] * [Group=Control]	-0.50	0.98	.61	-2.48	1.48
[App=with] * [Group=Control]	-0.51	1.26	.68	-3.04	2.01
[Moment_bin=.00] * [App=with] * [Group=Control]	0.40	1.33	.77	-2.30	3.10

Note. Moment_bin=.00 is a combination of the measure before and after the experiment.

Table 7
Estimates of Fixed Effects of the Subjective Valence Ratings

			95% Confidence Interval	
Parameter	Estimate	Sig.	Lower Bound	Upper Bound
Intercept	3.80	<.01	2.75	4.85
[Moment_bin=.00]	1.90	<.01	0.67	3.13
[App=with]	0.34	.62	-1.04	1.72
[Group=Control]	1.40	.07	-0.09	2.89
[Moment_bin=.00] * [App=with]	-0.54	.50	-2.16	1.07
[Moment_bin=.00] * [Group=Control]	-1.50	.09	-3.24	0.24
[App=with] * [Group=Control]	-0.34	.74	-2.37	1.69
[Moment_bin=.00] * [App=with] * [Group=Control]	1.54	.20	-0.83	3.92

Note. Moment_bin=.00 is a combination of the measure before and after the experiment.

3.3. Stressors

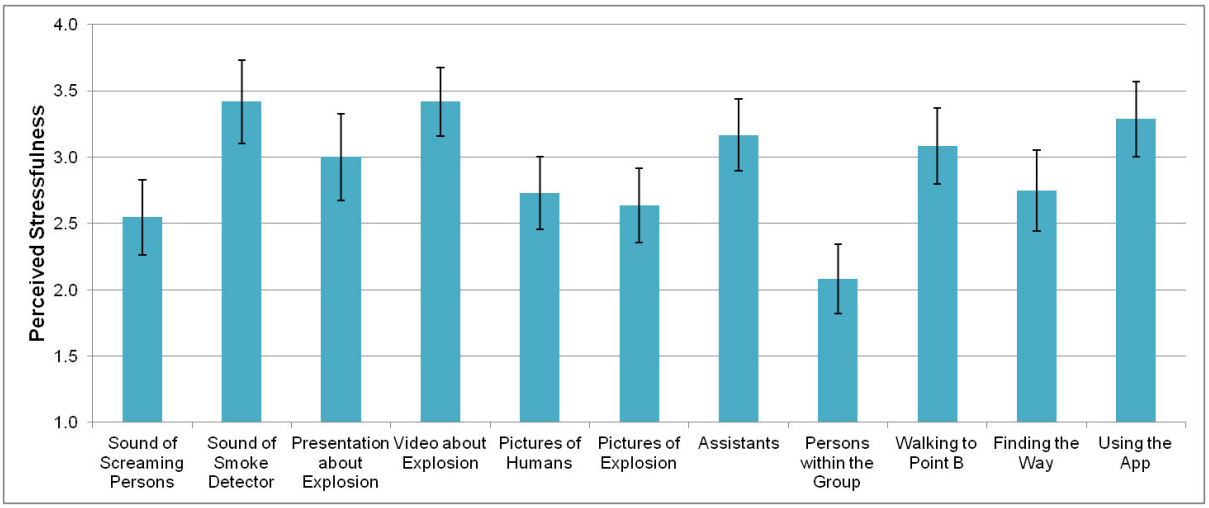


Figure 9. Mean ratings with error bars of the stressfulness ratings of elements in the experiment. 1 = not stressful to 5 = very stressful. Only the experimental group rated the elements. Most elements were intended to evoke stress (i.e., stressors) except for 'Pictures of Explosion', 'Presentation about Explosion', 'Video of Explosion', and 'Using the App'. N = 12 for all elements except for sound of screaming people, pictures of humans and pictures related to explosion, for them N = 11 due to missing ratings. Using the app was only rated by the participants with the app.

The people in the experimental group rated the stressfulness of the elements of the experiment. A score of one indicates 'not stressful' and five means 'very stressful'. The sound of the smoke detector (M = 3.42, SD = 1.08) and the video about the explosion (M = 3.42, SD = .90) were rated as most stressful. The figure shows that three more elements had a mean rating of three or higher. All other stressors or unintentional stressful elements were rated lower than three but higher than 2.5, except for the persons within the group (M = 2.08, SD = 0.90).

3.4. Other data

According to the data obtained through the smartphones and the observations of the assistants and confederates, all participants provided with the smartphone used the app for reporting and no participant seemed to be critically distracted by anything not related to the experiment. Moreover, the reported information about the incident indicated that all participants in the experimental group empathized with the scenario. The participants in the experimental group also behaved as they probably would during an emergency whereas in the control group participants did not. The assistants reported that the participants in the control group behaved calmly. One participant asked before the experiment whether they were supposed to run. They were instructed to behave however they felt like. All of them were walking slowly. Some of them were smiling and saying hello to the assistants, and they seemed to enjoy the experiment. One participant even started to laugh out loud when he saw a picture of happy people. Most participants of the experimental group ran or at least walked fast. They were more spread out over the path. Their facial expressions were more serious. A few of them who were ahead of the others behaved nervously and seemed to be very tense. Those participants did not

have the app. The participants in the experimental group arrived at point B individually, indicating that no groups were formed. In the control group, the participants arrived almost all at the same time.

The confederate in the experimental group reported to not have seen any of the posters because she was so focused on fleeing. Moreover, a participant in the experimental group indicated to have paid so much attention to the app that he missed some posters.

4. Discussion

First, the results are discussed. Then, limitations are discussed and suggestions for future research are presented.

4.1. Discussion of the Results

We start with the discussion of the subjective stress and SAM ratings, next the cortisol measures are discussed, thereafter other data, such as the observations, and finally the stressors.

4.1.1. Subjective ratings

The experimental group's subjective ratings were in favour of the success of the stress manipulation. The differences between the measures before to the measures during the experiment is similar to the one caused by a Stroop task but lower than caused by the Paced Auditory Serial Addition Task (PASAT) (Kolotylova et al., 2008). The mean stress ratings were a bit higher than the ratings in the Trier Social Stress Test (TSST), the MAST and the different versions of the cold pressure test (McRae et al., 2006; Schwabe et al., 2008; Smeets et al., 2012) and a bit lower than in the Sing-a-song stress test (SSST: (Brouwer & Hogervorst, 2014), the MMST, and the TSST for groups (TSST-G: Von Dawans et al., 2011). In the control group without the app, the mean rating during the experiment (M = 3.60, SD = 1.67) is higher than in the control group of the TSST-G (about M = 2.00). The control group with the app indicated similar stress ratings to the experimental group but higher dominance and valence ratings. The control group with and without the app showed higher ratings than in other studies. Perhaps, some participants were reminded of the Paris terror attacks of November 2015, which took place four days prior to the experiment, while watching the video and for this reason rated stress higher. However, this cannot be the only explanation, otherwise, the ratings in the control group with the app and without the app would be more alike.

The higher dominance ratings were perhaps related to the app, which could give them a feeling of control over the situation. Still, the high stress ratings in combination with relatively high valence and dominance ratings of some participants were contradictory, because they do not reflect a stress response. They may be explained by a testing effect. A testing effect in this case means, the participants with the app did not feel stress through the app itself but they felt more commitment to the experiment. The commitment could be expressed by filling in the questionnaire in a way that they think would help the experiment leader to confirm hypotheses or by imagining being in an emergency situation because they assumed that they are supposed to. They perhaps have assumed this since they

knew that an app for emergencies should be evaluated. This is called good-subject effect (Nichols & Maner, 2008).

The SAM ratings revealed that the control group was happier, less stressed, equally aroused and similar in perceived control. The expectations were that they felt more in control and less aroused than the experimental group. Consequently, the control condition should be redesigned to increase the feeling of control and to reduce the arousal.

4.1.2. Cortisol.

The cortisol measures did not confirm the hypothesis that the control group experienced no or only mild stress and the experimental group experienced stress. As expected the two groups cannot be compared directly because they were tested at different times of the day. That means, the cortisol levels of the two groups differed due to the circadian rhythm of cortisol causing the time of measurement to be a confound in this experiment. We assumed we could compare the tendency of the two groups. However, the tendencies did not reflect a stress response. This contradicts other stress protocols, such as TSST (Kirschbaum et al., 1993) or MMST (Reinhardt et al., 2012).

We will discuss some issues that are related to the time of day, the duration, the stressors, and the activities before the experiment to find an explanation for the unexpected findings. One may argue that the duration of the stress manipulation is too short to elicit cortisol. However, the MMST stress manipulation that has a similar duration as in our study was successful and a meta-study did not find an effect of duration on cortisol (Dickerson & Kemeny, 2004). The time of day for this study was determined based on the peak level in other studies. The measure was taken five minutes later than planned since several participants needed more time to fill out the second questionnaire. An explanation for the lack of results with cortisol is that within these five extra minutes the peak level passed, resulting in no significant increase of cortisol. Therefore, a professor in endocrinology at the university hospital in Würzburg (Hahner, 2016) was consulted to evaluate the moment of the second measure. She evaluated the time, 25 minutes after the experiment, as a good moment of measure that resembles the time lag used in clinical studies. Another issue is that the stressors evoke stress but not an elevation in cortisol. Literature about cortisol elicitation is contradicting. Some state that emotional stressors (Nejtek, 2002) and acoustic stressors (Bigert et al., 2005) are successful in eliciting cortisol, others state that this is not the case (Dickerson & Kemeny, 2004). Thus, the stressors in our experiment may have indeed been too weak to cause a cortisol reaction. Another issue is related to the activities before the experiment. The measures could be influenced by one of the activities (e.g. consuming food) that the participants should have refrained from before the experiment. We had to trust that the participants did indeed refrain from those activities. Another factor may have affected the findings negatively; namely, the number of extreme cortisol values. Either four participants suffered from unknown health issues or the results are due to a measurement error. Measurement errors seem more likely than four persons with unknown severe health issues. To sum up, measurement errors, the

selection of stressors, and that there was no control over the activities before the experiment may have been responsible for the lack of increase in cortisol that was found.

4.1.3. Other data

The results in 3.4. support the success of the scenario of an emergency situation. The reported information through the app, the interviews and the observation were in favour of the success of the simulation. The confederate in the experimental group had not been informed about what exactly would happen after leaving the first meeting room. The fact that the confederate failed to see the posters was probably due to attentional tunnelling. Thus, the confederate was stressed even though she had received some information about the experiment. The reported information through the app revealed unexpected findings. In the control group participants reported information objectively, however participants in the experimental group did not. The participants in the experimental group reported "(…) the one car exploded" "(…) resulted in killing of people, many of them got injured", "I think 2 people died on spot". Not reporting objectively is probably due to the impact of stress on information processing. Stress may impair among others the perception of stimuli (Ozel, 2001; Wickens et al., 2004) and decision making processes (Leach, 2004). This means that fewer stimuli are perceived to draw conclusions from due to attentional tunnelling and fewer alternatives are considered due to cognitive tunnelling (Wickens et al., 2004) when making a decision. This makes the report of wrong information more likely. The reported information also revealed that some participants reported an explosion and a fire; however, fire was meant to be a consequence of the explosion but not the incident itself. Participants indicated that their understanding of an explosion is related to fire. Future simulations of an explosion should take this into consideration.

4.1.4. Stressors

Most stressors were successful in evoking stress. The stressors 'Sound of screaming persons', 'Pictures of humans', and 'Persons within the group' got a mean rating below three, indicating low to moderate stressfulness. Thus, they were perceived as less stressful than expected. The sound of screaming people may have been rated lower than the sound of the smoke detector because it did not come in combination with a cognitive task. The participants namely did not start with answering questions in the app and looking for information. The study by Miki et al. (1998) found only an increase in cortisol when the acoustic stressor came in combination with an arithmetic task (cognitive task). The other people in the group may have been rated low on stressfulness since there was almost no interaction within the group; nobody needed the help of other participants, the distance between the most participants was too big to get influenced by other participants' behaviour and facial expressions, and some participants interacted more with the app than with their surroundings. If the group would be bigger and if some confederates would simulate the need for help, then the interaction among participants would probably increase, resulting in an effect of the stressor 'people in the crowd'.

The mean ratings of the presentation about the explosion, the video showing an explosion and using the app were not expected to indicate moderate stressfulness. Those elements were not meant as

a stressor but apparently contributed to stress. The presentation was meant to evoke some emotions but they were not assumed to be strong enough to evoke stress. Maybe the pictures of sad persons, the mentioning of death in the presentation, and the likelihood that people got severely hurt in the video worked as emotional stressors. Using the app may have increased the cognitive workload since some participants used the chat screen even when they did not feel able to. Improving the instructions about when to use this screen may reduce or eliminate the perceived stressfulness. The stressfulness of uncontrollability was not measured directly. However, the dominance ratings indicated that not all participants experienced a loss in control and a feeling of not being in control. This is an indication that the stressor was not successful for all persons.

To sum it up, the scenario was successful. Still, the visual stressors should be improved to evoke a stronger feeling of unhappiness and the stressor uncontrollability should be improved so that it affects more participants and has a stronger response. The stress manipulation is considered successful since the subjective ratings and the findings presented in 'other data' were in favour of the stress manipulation for the most participants, at least in the experimental group. The cortisol measure did not confirm the stress manipulation, probably because it was affected by measurement errors and the stressors were not strong enough for a cortisol reaction. Moreover, the high arousal and stress ratings in the control group with the app and the decrease in valence and dominance ratings in several participants in the control group did not support the success of simulation. However, the ratings are contradicting which indicates that they did not experience real stress. Thus, the simulation for the control group has to be improved, so that they are more relaxed during the experiment. This study showed that the simulation is applicable to evaluate the RESCUER App and, therefore, it is likely that the simulation can also be applied to other crowd-based rescue systems.

4.2. Limitations

Several limitations affected this study. Firstly, as mentioned the number of participants was small, even smaller than calculated with the g*-analysis. Thus, the number is too small to reflect a real crowd and behaviour in a crowd, such as altruistic behaviour (Pan, 2006). One requirement of the experiment was that it is culture-independent. However, the experiment did not allow to test whether culture affected the stress response. The number of representatives per culture was too small to compare their stress reaction and behaviour. Moreover, the representatives were not equally distributed over the 2x2 groups. Culture could influence the way participants rate questionnaires due to the tendency towards social desirability (He et al., 2015). Moreover, culture may affects the way participants deal with uncontrollability and unpredictability. According to Hofstede (1986), a cultural dimension is 'uncertainty avoidance', which "defines the extent to which people within a culture are made nervous by situations which they perceive as unstructured, unclear, or unpredictable" (Hofstede, 1986, p.308). Culture may have enhanced the individual differences of the stress response. Secondly, the experimental group had the impression that they know the path to point B due to an ambiguous

explanation. Therefore, the feeling of uncontrollability was probably low when leaving the meeting room. However, shortly after that when the path made an unexpected turn, the feeling of uncontrollability was probably higher than without the wrong impression. Therefore, the ambiguous explanation probably did not affect the experiment negatively in a big way. Thirdly, the experiment took place inside a building. The app is mainly meant for big industrial areas and large-scale events, which mostly include areas outside of buildings. In this study, limitations of the app were not unveiled that only appear in outside areas, such as poor contrast of the display due to sun. Fourthly, the simulated incident was described in the text, as taking place on the street and not in the building, which would had simulated the closeness to the incident better. This was done, since a simulation of an explosion directly in the room would have stressed the control group too much. Fifthly, the volume of the screaming people and the sounds in the control condition could not be measured. This should not have happened since it complicates the reproducibility of the simulation. In future simulations, the sounds should be played at least at a volume of 80 dB. Sixthly, in this simulation, it was assumed that running indicates that the participants felt the necessity to run and that this would resemble a real emergency. However, the assistants used the word 'hurry', which some participants maybe have interpreted as an instruction for running. Moreover, there are incidents where people do not feel the necessity to run. That means the argument that running is indicating the behaviour of a real emergency was poor. Seventhly, the other stressors in the questionnaire were not formulated well, e.g. walking to Point B did not express well that the activity of walking was addressed. Furthermore, an option saying 'don't know' should have been added to the questionnaire. Eighthly, the cortisol measures of the two groups were not at the same time of the day, which was a confound. Therefore, a direct comparison was not possible. Furthermore, some moderator variables were not assessed, such as personality traits and renal diseases. The subjective ratings were influenced by a testing effect, which is clear limitation. The subjective ratings were therefore less reliable. The testing effect probably occurred because the participants knew that the purpose of the experiment was the simulation of an emergency situation, that stress was probably evoked, and because they tried to be a good subject (Nichols & Maner, 2008). However, they did not indicate that one objective of the experiment was the induction of stress when asked about the purpose of the experiment.

4.3. Future Research

In future research the limitations of this study should be eliminated or their effect reduced or further investigated through a larger sample size, control for culture, improved objective and subjective measures, and a route that includes areas outside of buildings. The influence of culture on the stress ratings could be investigated by only two cultures which differ on several cultural dimensions, especially the dimension 'uncertainty avoidance' (e.g. Hofstede, 1986). Contrasting cultures would make the effect of culture more obvious. Alternatively, multiple experiments, each one with participants of a different culture, could be conducted. When an effect of culture is proved in future

studies, the effect could be used to increase the experienced stress for samples of the same culture or the effect could be eliminated to make the experiment culture independent, as it was supposed to be.

The sample size should be increased to better represent a crowd and crowd behaviour. For that, the location of the experiment has to be changed, since there are safety constraints in a building that is not meant for big crowds rushing through it. When selecting another location it is recommended to select a location that contains areas outside of buildings and that allows the participants to be observed better. The observations could serve to determine the behaviour of app users in comparison to non-users. It would be interesting to know what users look at. Do many users pay so much attention to the app that they miss important information in their surroundings? Are participants supporting each other when somebody needs help? Do app-users show any behaviour that would explain their high stress ratings? Do app users walk more slowly? Information like this would help to learn more about the influence of the app on behaviour and the validity of the experiment.

The simulation should be improved by intensifying the stress experience through an improved application of the stressors and an improved scenario. The improved simulation might evoke stress in a greater number of participants of the experimental group and an increased feeling of control and happiness in the control group. So far, we learned that the combination of acoustic stressors and high cognitive load is beneficial for evoking stress, and that app-users may fail to perceive visual stressors. Moreover, other people in a small group hardly contribute to the experience of stress. Still, the other people may contribute to an interaction effect with another stressor or they are effective as stressors when the number of persons is increased. Moreover, uncontrollability and emotional stressors should be improved. In future research, acoustic stressors could be added over a longer duration assuming that the combination of acoustic with cognitive stressors increases the stress. The acoustic stressor should be emotional, as well, to compensate for missed visual emotional stressors. Emotional pictures that have a stronger affect on emotions should be selected. The IAPS could be addressed for pictures (Lang et al., 1997). The assistants should be located in less obvious spots, so that the participants only know where to go very shortly before a turn. Moreover, the incident could be simulated to be closer to the participants in order to increase the stress experience. Future research should address, for instance, whether a larger number of persons simulating a crowd functions as a stressor and whether the presentation about the explosion contributes even more to stress when it includes acoustic and emotional stressors. Even if people do not walk fast or run in all types of emergency, we advice for the evaluation of a crowd-based rescue system to stimulate the people to walk or run so that the most extreme use context is tested.

The testing effect should be eliminated. This study showed that only instructing the participants to accept the experiment as it is even if it does not make sense to them (see appendix F for the instruction) is not enough. The study about the good-subject effect found, that “participants reporting more positive attitudes [towards the experiment and experiment leader] were more likely to behave as a good subject” (Nichols & Maner, 2008, p. 161). Some participants were friends of the experiment

leader and others might have had positive attitudes because they were master students, as well. Not informing the participants that the experiment leader is a student, who is conducting the experiment as part of her master thesis, may reduce the tendency to 'help' the experiment leader to confirm expectations. Furthermore, the scenario of the control group should be improved so that it is telling a coherent story with fewer aspects that make no sense to the participants; for instance, the posters with happy people. Then, the participants probably do not try to interpret the situation in a way that is consistent with prior knowledge about the study resulting in more reliable subjective ratings. Moreover, the feeling of control should be increased in the control group and arousal decreased. The control group should receive even more information about what is going to happen, for instance in form of photos and an adjustment of the text describing the scenario. They should be reminded during the experiment that they are not judged for what they report in order to reduce the feeling of uncontrollability and arousal. Furthermore, assistants could walk with the participants so that they always have a contact person in case questions arise. After the experiment, some participants of the control groups should be interviewed. In case the participant indicated that they felt stressed, the interview will reveal whether they indeed felt stressed or whether a testing effect occurred.

The cortisol measure for the experimental and control group should be taken at exactly the same time of the day. This should improve the comparability of the two groups. Moreover, blood samples instead of saliva samples could be used to precisely determine the cortisol levels since in our study taking the sample took a relative long amount of time (ca. 3 min) and measurement errors occurred. Medical staff should check the participants for symptoms of lack of cortisol and the participants should fill in a questionnaire, for instance the Brief Symptom Inventory (BSI: Derogatis & Melisaratos, 1983), to control for psychological disorders.

A continuous measure of stress would reveal the stress levels at the different zones. A continuous measure is difficult to obtain since the participants are moving. Some studies obtained continuous measures, for example, of heart rate (Duncko et al., 2009) and skin conductance (Gomez & Danuser, 2004), but in those studies, the participants were not walking. An idea is to obtain a baseline to compare the heart rate and skin conductance to the measurement during a stress experience. However, in that case, the participants have to walk during the experiment at the same pace as during the baseline, which is difficult to accomplish. So far, it is possible to detect a startle response using skin conductance even at six km/h (Schumm et al., 2008). Maybe in the future, the devices and the analyses develop further, so that the reaction to a longer lasting stressor can be distinguished from the reaction to movement and other processes, such as attention (Dawson, Schell, Filion, & Berntson, 2012).

The simulation can be conducted with an improved version of the app. This will reveal how much the newer version has improved compared to the former one. Future research should also test whether the simulation is successful when participants imitating rescue workers walk the other way around, namely from the safe place to the incident, while using the information provided by the crowd.

5. Conclusion

While planning an evaluation of the RESCUER App, the need for a simulation of an emergency situation for crowd-based rescue systems arose. Therefore, a simulation was created taking the use context of a crowd-based rescue system into consideration and the induction of stress. The stressors were selected on the basis of literature and the suitability for emergency situations. We tested whether the simulation is suitable to evaluate a crowd-based rescue system by evaluating the RESCUER App. The results indicated that the simulation is successful but could be further improved. The results are an indication that the simulation is also suitable for other crowd-based rescue systems. The simulation may have to be adjusted to the specifications of other systems and the incident type. The incident type could be adapted by changing the scenario, which is modelled by the video, text, and pictures. Future research should address whether the simulation is also suitable to evaluate a crowd-based rescue system while moving from a safe place to the incident.

References

Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*, *34*(10), 1423–1436. http://doi.org/10.1016/j.psyneuen.2009.06.011

Allen, A. P., Kennedy, P. J., Cryan, J. F., Dinan, T. G., & Clarke, G. (2014). Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. *Neuroscience and Biobehavioral Reviews*, *38*, 94–124. http://doi.org/10.1016/j.neubiorev.2013.11.005

Bahr, H. J., Betz, C., Brenner, K., Eikelberg, T., Hohausen, B., Junghans, S., … Wulftange, B. (2015). *Öffentliche Veranstaltungen*. Merching: Forum Verlag Herkert GMBH.

Bergstrand, F., & Landgren, J. (2009). Information Sharing Using Live Video in Emergency Response Work. In *Proceedings of the 6th International ISCRAM Conference*.

Bigert, C., Bluhm, G., & Theorell, T. (2005). Saliva cortisol - A new approach in noise research to study stress effects. *International Journal of Hygiene and Environmental Health*, *208*(3), 227–230. http://doi.org/10.1016/j.ijheh.2005.01.014

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*. Retrieved from http://dionysus.psych.wisc.edu/methods/Stim/ANEW/ANEW.pdf\nhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Affective+Norms+for+English+Words+(+ANEW+):+Instruction+Manual+and+Affective+Ratings#0\nhttp://scholar.google.com/scholar?hl=en&btnG=S

Brouwer, A.-f., & Hogervorst, M. A. (2014). A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST). *Frontiers in Neuroscience*, *8*(July), 1–8. http://doi.org/10.3389/fnins.2014.00224

Carifio, J., & Perla, R. J. (2007). Ten Common Misunderstandings , Misconceptions , Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, *3*(3), 106–116.

Croy, I., Olgun, S., & Joraschky, P. (2011). Basic emotions elicited by odors and pictures. *Emotion*, *11*(6), 1331–1335. http://doi.org/10.1037/a0024437

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*(2), 468–477. http://doi.org/10.3758/s13428-011-0064-1

Dawson, M. E., Schell, A. M., Filion, D. L., & Berntson, G. G. (2012). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 157–181). Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511546396.007

Delignières, D., Brisswalter, J., & Legros, P. (1994). Influence of Physical Exercise on Choice Reaction Time in Sports Experts : the Mediating Role of Resource Allocation. *Journal of Human Movement Studies*, *27*, 173–188.

Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory: an introductory report. *Psychological medicine*, *13*(03), 595-605. http://dx.doi.org/10.1017/S0033291700048017

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*(3), 355–

391. http://doi.org/10.1037/0033-2909.130.3.355

Duncko, R., Johnson, L., Merikangas, K., & Grillon, C. (2009). Working memory performance after acute exposure to the cold pressor stress in healthy volunteers. *Neurobiology of Learning and Memory*, *91*(4), 377–381. http://doi.org/10.1016/j.nlm.2009.01.006

Fernández, C., Pascual, J. C., Soler, J., Elices, M., Portella, M. J., & Fernández-Abascal, E. (2012). Physiological responses induced by emotion-eliciting films. *Applied Psychophysiology Biofeedback*, *37*(2), 73–79. http://doi.org/10.1007/s10484-012-9180-7

Fokkinga, S. F., & Desmet, P. M. a. (2013). Ten ways to design for disgust, sadness, and other enjoyments: A Design Approach to Enrich Product Experiences with Negative Emotions. *International Journal of Design*, *7*(1), 19–36.

Gathmann, B., Schulte, F. P., Maderwald, S., Pawlikowski, M., Starcke, K., Schäfer, L. C., … Brand, M. (2014). Stress and decision making: neural correlates of the interaction between stress, executive functions, and decision making under risk. *Experimental Brain Research*, *232*, 957–973. http://doi.org/10.1007/s00221-013-3808-6

Glendon, a. I. (2011). *Neuroscience and young drivers*. *Handbook of Traffic Psychology*. Elsevier. http://doi.org/10.1016/B978-0-12-381984-0.10009-8

Gohm, C. L., Corser, G. C., & Dalsky, D. J. (2005). Emotional intelligence under stress: Useful, unnecessary, or irrelevant? *Personality and Individual Differences*, *39*(6), 1017–1028. http://doi.org/10.1016/j.paid.2005.03.018

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology*, *53*(2), 91–103. http://doi.org/10.1016/j.ijpsycho.2004.02.002

Hahner, S. , (personal communication, January 2016)

He, J., van de Vijver, F. J. R., Dominguez Espinosa, A., Abubakar, A., Dimitrova, R., Adams, B. G., … Villieux, A. (2015). Socially Desirable Responding: Enhancement and Denial in 20 Countries. *Cross-Cultural Research*, *49*(3), 227–249. http://doi.org/10.1177/1069397114552781

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, *34*(2), 163–171. http://doi.org/10.1016/j.psyneuen.2008.10.026

Hellhammer, J., & Schubert, M. (2012). The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology*, *37*(1), 119–124. http://doi.org/10.1016/j.psyneuen.2011.05.012

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, *92*, 84–89. http://doi.org/10.1007/s00421-004-1055-z

Hobfoll, S. E. (1989). Conservation of resources. A new attempt at conceptualizing stress. *The American Psychologist*, *44*(3), 513–524. http://doi.org/10.1037/0003-066X.44.3.513

Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations*, *10*, 301–320. http://doi.org/10.1016/0147-1767(86)90015-5

Kalat, J. W. (2008). *Biological Psychology* (10th ed.). Belmont: Wadsworth.

Kemeny, M. E. (2003). The Psychobiology of Stress. *Current Directions in Psychological Science*, *12*(4), 124–129. http://doi.org/10.1111/1467-8721.01246

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The "Trier Social Stress Test"--a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*. http://doi.org/119004

Kolotylova, T., Koschke, M., Ebner-priemer, U., Kleindienst, N., Bohus, M., Bär, K.-J. Ü. R., & Schmahl, C. (2008). Entwicklung des „ Mannheimer Multikom ponenten- Stress-Test " (MMST). *Psychotherapie, Psychosomatik, Medizinische Psychologie*, *60*, 64–72. http://doi.org/10.1055/s-0028-1103297

Kudielka, B. M., Hellhammer, D. H., & Kirschbaum, C. (2007). Ten years of Research with the Trier Social Stress Test (TSST)- Revisited. In *Social Neuroscience* (pp. 56–83). London: The Guilfort Press.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*. *Psychology*.

Leach, J. (2004). Why people "freeze" in an emergency: Temporal and cognitive constraints on survival responses. *Aviation Space and Environmental Medicine*, *75*(6), 539–542.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., … Choudhury, T. (2012). StressSense: Detecting Stress in Unconstrained Acousitc Environments using Smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12* (p. 351). http://doi.org/10.1145/2370216.2370270

Mason, J. W. (1968). A review of psychoendocrine research on the pituitary-Adrenal Cortical System. *Psychosomatic Medicine*, *30*(5), 576–607. http://doi.org/1968/09000

McRae, A. L., Saladin, M. E., Brady, K. T., Upadhyaya, H., Back, S. E., & Timmerman, M. A. (2006). Stress reactivity: biological and subjective responses to the cold pressor and Trier Social stressors. *Human Psychopharmacology: Clinical and Experimental*, *21*(6), 377–385. http://doi.org/10.1002/hup.778

Mendl, M. (1999). Performing under pressure: Stress and cognitive function. *Applied Animal Behaviour Science*, *65*(3), 221–244. http://doi.org/10.1016/S0168-1591(99)00088-X

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the international affective picture system. *Behavior Research Methods*, *37*(4), 626–630. http://doi.org/10.3758/BF03192732

Miki, K., Kawamorita, K., Araga, Y., Musha, T., & Sudo, A. (1998). Urinary and salivary stress hormone levels while performing arithmetic calculation in a noisy environment. *Industrial Health*, *36*, 66–69. http://doi.org/10.2486/indhealth.36.66

Mogg, K., Bradley, B. P., & Hallowell, N. (1994). Attentional bias to threat: roles of trait anxiety, stressful events, and awareness. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *47*(4), 841–864. http://doi.org/10.1080/14640749408401099

Nakagawa, E. Y., Soares Santos, D., Bueno Ruas de Oliveira, L., & Duran, A. (2015). *Deliverable 5.2.1 Evaluation Report of the Mobile Crowdsourcing Solution 1*. Retrieved from http://www.rescuer-project.org/?page_id=11037

Nass, C., Jung, J., Groen, E. C., & Villela, K. (2015b). Talk to me, crowd! Interaction Modes for Emergency Mobile Apps. *Not yet Published*.

Nass, C., Magin, D., Breiner, K., Villela, K., Groen, E., & Jung, J. (2015). Conceptual Model of Mobile User Interaction in Emergencies 3, Deliverable 2.1.3.

Nejtek, V. a. (2002). High and low emotion events influence emotional stress perceptions and are associated with salivary cortisol response changes in a consecutive stress paradigm. *Psychoneuroendocrinology*, *27*(3), 337–352. http://doi.org/10.1016/S0306-4530(01)00055-5

Nichols, A. L., & Maner, J. K. (2008). The Good-Subject Effect: Investigating Participant Demand Characteristics. *The Journal of General Psychology*, *135*(2), 151–166. http://doi.org/10.3200/GENP.135.2.151-166

Niculescu, A., Cao, Y., & Nijholt, A. (2008). Stress and Cognitive Load in Multimodal Conversational Interactions. *International Conference on Human-Computer Interaction*.

Niculescu, A., Cao, Y., & Nijholt, A. (2010). Manipulating Stress and Cognitive Load in Conversational Interactions with a Multimodal System for Crisis Management Support. In *Lecture Notes in Computer Science* (pp. 134–147). http://doi.org/10.1007/978-3-642-12397-9_11

Oei, N. Y. L., Everaerd, W. T. a M., Elzinga, B. M., van Well, S., & Bermond, B. (2006). Psychosocial stress impairs working memory at high loads: An association with cortisol levels and memory retrieval. *Stress*, *9*(3), 133–141. http://doi.org/10.1080/10253890600965773

Ozel, F. (2001). Time pressure and stress as a factor during emergency egress. *Safety Science*, *38*, 95–107. http://doi.org/10.1016/S0925-7535(00)00061-8

Pan, X. (2006). *Computational modeling of human and social behaviors for emergency egress analysis*. Retrieved from http://eil.stanford.edu/egress/publications/PhD_Thesis_Pan_CIFE.pdf

Penley, J. A., & Tomaka, J. (2002). Associations among the Big Five, emotional responses, and coping with acute stress. *Personality and Individual Differences*, *32*(7), 1215–1228. http://doi.org/10.1016/S0191-8869(01)00087-3

Quarantelli, E. L. (1975). Panic behavior: some empirical observations. In *American Institute of Architects Conference on Human Response to Tall Buildings, July 19, 1975, Chicago, Illinois*.

Raff, H., & Trivedi, H. (2013). Circadian rhythm of salivary cortisol, plasma cortisol, and plasma ACTH in end-stage renal disease. *Endocrine Connections*, *2*(1), 23–31. http://doi.org/10.1530/EC-12-0058

Reinhardt, T., Schmahl, C., Wüst, S., & Bohus, M. (2012). Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the Mannheim Multicomponent Stress Test (MMST). *Psychiatry Research*, *198*, 106–111. http://doi.org/10.1016/j.psychres.2011.12.009

Ribeiro, R. L., Pompéia, S., & Amodeo Bueno, O. F. (2005). Comparison of Brazilian and American norms for the International Affective Picture System (IAPS). *Revista Brasileira de Psiquiatria*, *27*(3), 208–215. http://doi.org/10.1590/S1516-44462005000300009

Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, *53*, 61–97. http://doi.org/DOI: 10.1016/0001-6918(83)90016-1

Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (2014). ANGST: Affective

norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods*, *46*, 1108–1118. http://doi.org/10.3758/s13428-013-0426-y

Schultheis, M. T., & Manning, K. J. (2011). *Neuroscience and older drivers*. *Handbook of Traffic Psychology*. Elsevier. http://doi.org/10.1016/B978-0-12-381984-0.10010-4

Schumm, J., Bächlin, M., Setz, C., Arnrich, B., Roggen, D., & Tröster, G. (2008). Effect of movements on the electrodermal response after a startle event. In *Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare 2008, PervasiveHealth* (pp. 315–318). http://doi.org/10.1109/PCTHEALTH.2008.4571101

Schwabe, L., Haddad, L., & Schachinger, H. (2008). HPA axis activation by a socially evaluated cold-pressor test. *Psychoneuroendocrinology*, *33*(6), 890–895. http://doi.org/10.1016/j.psyneuen.2008.03.001

Schwabe, L., & Wolf, O. T. (2010). Emotional modulation of the attentional blink: is there an effect of stress? *Emotion*, *10*(2), 283–288. http://doi.org/10.1037/a0017751

Selye, H. (1976). Forty years of stress research: principal remaining problems and misconceptions. *Canadian Medical Association Journal*, *115*(1), 53–56.

Smeets, T., Cornelisse, S., Quaedflieg, C. W. E. M., Meyer, T., Jelicic, M., & Merckelbach, H. (2012). Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology*, *37*(12), 1998–2008. http://doi.org/10.1016/j.psyneuen.2012.04.012

Srygley, J. M., Mirelman, A., Herman, T., Giladi, N., & Hausdorff, J. M. (2009). When does walking alter thinking? Age and task associated findings. *Brain Research*, *1253*, 92–99. http://doi.org/10.1016/j.brainres.2008.11.067

Stankovic, A., Fairchild, G., Aitken, M. R. F., & Clark, L. (2014). Effects of psychosocial stress on psychophysiological activity during risky decision-making in male adolescents. *International Journal of Psychophysiology*, *93*(1), 22–29. http://doi.org/10.1016/j.ijpsycho.2013.11.001

Stevenson, R. a, & James, T. W. (2008). Affective auditory stimuli: characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories. *Behavior Research Methods*, *40*(1), 315–321. http://doi.org/10.3758/BRM.40.1.315

Streefkerk, J. W. (2011). *Doing the right task*. Technische Universiteit Delft.

Takai, N., Yamaguchi, M., Aragaki, T., Eto, K., Uchihashi, K., & Nishikawa, Y. (2004). Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults. *Archives of Oral Biology*, *49*(12), 963–968. http://doi.org/10.1016/j.archoralbio.2004.06.007

Vollrath, M. (2001). Personality and stress. *Scandinavian Journal of Psychology*, *42*(4), 335–347. http://doi.org/10.1111/1467-9450.00245

von Dawans, B., Kirschbaum, C., & Heinrichs, M. (2011). The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology*, *36*(4), 514–522. http://doi.org/10.1016/j.psyneuen.2010.08.004

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. E. (2004). *An Introduction to Human Factors Engineering* (2nd ed.). New Jersey: Pearson Education International.

Yehuda, R., Teicher, M. H., Trestman, R. L., Levengood, R. A., & Siever, L. J. (1996).

Cortisol regulation in posttraumatic stress disorder and major depression: A chronobiological analysis. *Biological Psychiatry*, *40*(2), 79–88. http://doi.org/10.1016/0006-3223(95)00451-3

Music

Secession studios (2014), *Ash*, from the album *Epic Classics*

YouTube (2013), *Sendung mit der Maus Intro Extended,* Published at 12.04.2013 on https://www.youtube.com/watch?v=tLpdMToxngU

Explosions in the Sky (2004), *Home*, from the album *Friday Night Lights*

Websites

CSEA, *ANET*, attended at 23-02-2016, http://csea.phhp.ufl.edu/media/anetmessage.html

RESCUER, derived at 11-02-2016 from http://www.rescuer-project.org/wordpress/wp-content/uploads/2015/01/RESCUER_FLYER.pdf

YouTube (2015), attended at 12-11-2015, *Intense Gas Explosion Caught on Camera in South Korea July 12th 2013 Massive,* https://www.youtube.com/watch?v=yJTUxr2zOxk
(The link is not active any longer but searching for the title leads to the same video. Alternative Link: https://www.youtube.com/watch?v=rT8KvgGDwrg&nohtml5=False)

Appendix A. Laboratory Stress Protocols

The table gives an overview of already existing protocols and commonly used stressors.

Name	*Procedure*	*Control Condition*	*Measuring Stress*	*Stressors*	*Ethical Constraints*	*Reference*
Trier Social Stress Test (TSST)	Simulated job interview (10 min preparation for a 5 min speech in front of 3 'managers'; video and voice analysis of test person's performance was announced but not conducted, 'managers' asked mental arithmetic tasks)	Same blood/ saliva measurements but no test; or friendly version of TSST (Wiemers, Schoofs, & Wolf, 2013)	Cortisol (Blood or saliva) , in first experiment also heart rate, ACTH, GH, prolactin	Audience, mental arithmetic, public speaking, anticipatory period	Nothing mentioned	(Kirschbaum et al., 1993)
Maastricht acute stress test (MAST)	Combines TSST and (SE)CPT: 5× trials (hand in 2 C cold water), telling participant that duration of trial randomly choosen by computer (max. 90s) seconds while being watched and video-tapped, between trials mental arithmetic task, when mistake than negative feedback and starting again, participants are told that		Cortisol, blood pressure, subjective ratings, Salivary alpha-amylese	Social evaluative, Uncontrollability, Unpredictability, mental arithmetic, thermal/pain	Approved by ethics committee of faculty of PSY and neuroscience (University of Maastricht)	(Smeets et al., 2012)

	duration of arithm. Task is determined by computer (in real all durations were fixed and the same for all participants)					
Cold Pressure test (CPT)	putting one hand into 0-4 C cold water for as long as possible with a maximum of 3 minutes	35-37 C warm water	Blood pressure, ECG, cortisol, subjective	Thermal/pain		(Schwabe et al., 2008)
Social evaluative cold pressure test (SECPT)	CPT + being watched by an experimenter of opposite gender + video recording for facial analysis	Same procedure but 35-37 C war m water	Blood pressure, ECG, cortisol, subjective	Thermal/pain, social evaluation		(Schwabe et al., 2008)
Prolonged SECPT (P-SECPT)	SECPT but putting hand in cold water multiple times	Same procedure but 35-37 C war m water		Thermal/pain, social evaluation		(Smeets et al., 2012) //no paper found about original test invention
Imaging MAST (i-MAST)	MAST for fMRI, Instead of cold water a thermode is used at the forearm	none	Subjective, cortisol, alpha-amylase	Social evaluative, Uncontrollability, Unpredictability, mental arithmetic, thermal/ pain	Approved by ethics committee of faculty of PSY and neuroscience	(Quaedflieg, Meyer, & Smeets, 2013)

					(University of Maastricht)	
TSST for Groups (TSST-G)	TSST with more participants at the same time	Protocol without social evaluation and motivated performance	Heart rate, cortisol, subjective	Audience, mental arithmetic, public speaking, anticipatory period		(von Dawans et al., 2011)
Sing-a-Song Stress Test (SSST)	Singing a song while being watched by two confederates	none	skin conductance, heart rate	Social-evaluative	In accordance with the declaration of Helsinki ; approved by local ethic committee	(Brouwer & Hogervorst, 2014)
Mannheim Multicomponent stress test (MMST)	Paced auditory serial addition task (PASAT),		Hear rate variability, subjective, cortisol	Cognitive, emotional, motivational, acoustic	Local ethic committee	(Kolotylova et al., 2008) (Reinhardt et al., 2012)

Quaedflieg, C. W. E. M., Meyer, T., & Smeets, T. (2013). The imaging Maastricht Acute Stress Test (iMAST): A neuroimaging compatible psychophysiological stressor. *Psychophysiology*, *50*(8), 758–766. http://doi.org/10.1111/psyp.12058

Wiemers, U. S., Schoofs, D., & Wolf, O. T. (2013). A friendly version of the Trier Social Stress Test does not activate the HPA axis in healthy men and women. *Stress*, *16*(2), 254–260. http://doi.org/10.3109/10253890.2012.714427

Appendix B. g*-Analysis

A 'g*-analysis' was conducted to get to know how many participants are needed. We expect a large effect size since in other studies about stress the effect size was large as well. The analysis was conducted several times, since the number of groups and measures are different depending what will be analysed.

Two groups:

- Control group
- Experimental group

Four groups:

- Control group with app
- Control group without app
- Experimental group with app
- Experimental group without app

Two measures:

- Cortisol level

Three measures:

- Subjective stress ratings
- SAM ratings

ANOVA: repeated measures, within-factor

Effect size f	0.4 (large)			
Number of groups	2	2	4	4
Number of measures	2	3	2	3
Total sample size	24	18	24	20

The chosen number of participants is 24. That means 12 in the experimental group and 12 in the control group.

Appendix C. Protocol of the Experiment

	Name step	*Approx. Duration in Minutes*	*place*	*content*	*Purpose/ description*
1	welcome	5	Meeting room	Welcoming participants,	The participants enter the meeting room. Enough time is planned for this so that the participants do not feel stressed before the first stress measures (=baseline).
2	Participant number	2-4	Meeting room	Assigning participant number	The participant get a participant number. This activity and handing out the smartphones may seem more appropriate after the "explanation experiment" but it is done at this point of time so that people who are a bit late can still join the experiment. The participant number is written on a piece of tape that is sticked to the tables that offered a good view on the screen. The participants are asked to take the tape with them by putting it in their pocket or by sticking it to their clothes.
3	Handout smartphones		Meeting room	Randomly assign smartphone	The participants draw a small piece of paper. When the paper says "Smartphone" then they get one and when it says "NoSmartphone" they do not get one. Since, there are 24 participants + two confederate in total. In the experimental, as well as, in the control group, there are six or seven persons with a smartphone and six or seven without depending on the availability of phones.
4	Explanation about psy. Experiments	3	Meeting room	Showing a few slides	It is likely that some participants have never participated in a psychological experiment before. Therefore, a short explanation is given about psy. Experiments in general. They are informed that it is important to be

	In general				themselves, thus, to answer honestly, that means, not socially desirable and not fitting the purpose (at least what they think is the purpose) of the experiment. Furthermore, they are told that may not all elements of the experiment seem reasonable to them and that they just should accept the experiment like it is. They are also informed about potential distress and stress caused by the experiment and their right to withdraw from their participation in the experiment at any point in time. By giving them an introduction to psy. experiment the participants feel more in control and therefore more comfortable.
5	Informed consent	3	Meeting room	Signing informed consent	The participants are given the informed consent and are asked to sign it. The informed consent is in line with the requirements asked by Fraunhofer and the ones of international standards for psychological research.
6	Cortisol measure1	2	Meeting room		The participants get a short introduction about how to give the saliva sample. After that they get a bit of water to clean their mouth. Then, the saliva sample is collected. They are asked to write their participant number on the sticker of the sample equipment.
7	Questionnaire1	6	Meeting room	Filling in the questionnaire	The questionnaire asks for the participant's current stress level (Likert scale) and emotional state (SAM). Those measures are necessary as a baseline for the later measures. Furthermore, the participants are asked about variables that may moderate the stress effect such as experience with emergencies. Moreover demographic data are reported.
8	Explanation experiment	4	Meeting room	Introduction to RESCUER app, Introducing the	1. Introduction to RESCUER App: They are introduced to the RESCUER app (purpose, icon, different screens). It would be unrealistic to throw the participants into the deep end with regard to the app since if they would decide

				tasks	to install the app on their smartphone, they would also know about the purpose of the app, how the icon of the app looks like, and hopefully they also have checked out the demonstration of the app. 2. Task description: the participants are informed about the two task they have to perform later. A short explanation is given about when they have to perform the tasks. They will namely see first a scenario consisting out of a text and a video, then they see a slide saying 'GO'. This slide will be the start sign for the tasks. First task: go to point B. Both groups were informed that the meeting room is the place where the incident will take place and that point B is the safe place. They were also informed about the interaction modes appropriate to those areas. They are free to use all modes when they like to. They are shown the position of point B on a map. In the map for the control group, also the way how to get there is depicted. It is not the most direct way so that they need more time to get there. In a real situation you also need some time to leave the area of risk. The experimental group does not receive information about the way. This is done with the purpose of creating a feeling of uncontrollability. Task 2: Both groups are told that they might see more things on their way to point B that may are of interest to the rescue workers. Those information and the information they get out of a video should be reported to the command centre. The participants with the app are asked to report the information on their way to point B and the one without a smartphone are told that they will receive a questionnaire later on. After both tasks were explained, the participants are asked to describe themselves what they have to do. This has the purpose of ensuring that

					everyone understood the tasks. They also have the possibility to ask questions. A pre-test had yield that the tasks were not clear enough. The participants in the pre-test reported that they knew what to do but as soon as they left the room, they forgot. A poster with the slides representing the tasks with pictures hang on the way to point B as a reminder.
9	Presentation	4	Meeting room	Watching a presentation	The presentations had the purpose of bringing the participants in the right mind set. The presentations show pictures and short descriptive sentences. There are 9 slides with photos, 4 of them depicting humans, 3 of them explosions. Both presentations last for 2:22 minutes.
				Experimental group: serious presentation about the firework disaster in Enschede	The slides have a black background with white letters on it. The presentation is accompanied by instrumental music (sad & dramatic: https://www.youtube.com/watch?v=BeEDQMLtyGY). The presentation introduces the watcher to the firework explosion in Enschede in the year 2000. As a consequence of the firework explosion, several houses burnt down totally. This part of the firework disaster is neither depicted in photos nor described by text. The reason for this is that the participants should only deal with the topic of explosion and not with fire. The goal of the presentation is it to demonstrate to the participants that explosions are realistic and that they can happen even close to them. Moreover, the presentation is supposed to create a negatively emotional state. In this state, the participants are supposed to better empathize with the scenario described by the text and video.
				Control group: funny presentation	White background with black letters. The presentation is accompanied by cheerful music (https://www.youtube.com/watch?v=tLpdMToxngU). The control group is also watching a presentation to keep the design of the

				about the Mythbusters letting explode things	experiment in the two groups as similar as possible. Furthermore, creating a funny context is supposed to work as a protective frame (Fokkinga & Desmet, 2013). The effect is comparable with funny homevideos that show people getting hurt even so watchers can laugh about it. The presentation tells a story about the Mythbusters, a American TV program. The Mythbusters team tests myth in a scientific but entertaining manner. The explosions created by Mythbusters take place under controlled conditions and are obviously enjoyed by them. The presentation shows photos of explosions created by the Mythbusters. The photos, depicting humans, show smiling faces of the team members. Other photos show an exploding car. Thus, watchers can enjoy looking at the explosions because they know nothing bad will happen.
10	Scenario description		Meeting room	Reminding about the scenario	The participants are reminded of the process that has already been explained when introducing the tasks. In the experimental group, the reminder is presented in a hectic manner to continue the 'mind set' that already started with the presentation.
11	Text	3	Meeting room	The text describes a scenario with which the participants should empathize	The text is presented on power point slides. The text describes a situation that takes place in the meeting room with the purpose that the participants can easily empathize with the scenario. The same music as in the presentation is played. The text is included in a movie that also contains the video. The reason for this is that then the presentation of the text is equally long and that the video is played directly after the text without any interruption.
				Experimental:	The text describes a state of uncontrollability by describing a state of negative foreboding. Some negative words were bold print to catch the attention of the participants. The text gives no clue about what will happen in the video.

				Control:	The text describes a scenario which says that the participants are in a safe distance to the incident. The incident, an explosion, was caused by the Mythbusters doing something wrong. The scenario should follow-up the safe and funny context of the presentation. The text says that something with the experiment of the Mythbusters went wrong. Neutral words are bold print to make the layout of the text alike the one of the experimental group.
12	Video	1	Meeting room		This video is retrieved from YouTube: ("Intense Gas Explosion Caught on Camera in South Korea July 12th 2013 Massive" https://www.youtube.com/watch?v=yJTUxr2zOxk). The video shows an explosion in a street next to two men. After the explosion one man is running away. The watcher can assume that the other man may is injured but there is no injured person recorded on the video. This video is chosen because (1) there is almost no fire depicted, (2) it is not showing any injured person nor faces reflecting shock/pain/fear, (3) it presents a situation easy to empathize with since it takes place in a normal street. The selection criteria are intended to guarantee the suitability to both groups and to ensure that the participants are reporting an explosion and not a fire. About (2): showing injured persons or faces would be a stimuli evoking negative emotions. However, the control group should not be affected by negative stimuli. Furthermore, according to a former study, individuals under stress reported injured persons even when there were no injured person depicted. In the meanwhile the question about injured persons was reworded. We want to test whether they are still reporting injured persons when there are none depicted.
13	A to B		Corridor	Getting from	Participants without smartphone: They are told that they are later asked to

			at Fraunhofer	meeting room to meeting point B while reporting the incident	report the incident and therefore, also should pay attention to their surroundings Experimental: Exposure to sound of fire alarm (ca. 80dB). The loudness was not varied as suggested by (Kolotylova et al., 2008). The exposure to the sound in our experiment is too short for habituation effects to occur. There are pictures on the way, printed in size xx. The pictures show damaged houses and are meant to be reported by the participants. Furthermore, there are pictures of faces of a crying child and of people fleeing to keep the emotional stressor high. Control: Exposure to slow instrumental music (ca. 50 dB) They see the same pictures of damaged houses and pictures of smiling people. In both groups also neutral photos (houses, street) are shown because in a real emergency you also see neutral stimuli. Moreover, only showing relevant photos will tell the participants that every photo is of importance. By presenting neutral and relevant photos they have to decide which photos should be reported.
14	Chat function		Point B		The participants are now in the cold zone. They have some time to finish sending information via the chat function.
15	Questionnaire 2	6	Point B		Questions about stress level and emotional state have to be answered for two different points in time. One time for how they felt during the experiment and one time how they feel at the moment. The participants were not asked to indicate their stress level directly during the experiment because this would

					have disturbed the flow of the experiment. The questionnaire also contains open questions about the app. In the open questions the screenshots of the different screen of the app are combined with questions about the screen. The questions were about what the participant like and dislike about the screen and about any other comment about the screen. Participants without smartphone: They are asked the same questions about the incident 'explosion' than provided at the app. Instead of the chat function, they have the possibility to write down anything else about the incident they like to report imagine reporting the incident to a command centre. The questions are provided in written form to keep the medium of the report as similar as possible. The experimental group was also asked how stressful they experience the elements of the experiment such as the sound of fire alarm.
16	Point B	2	Point B	Asking back smartphones	The participants have to stick the tape with their participant number to the phone.
17	Cortisol measure 2	2	Point B		20 minutes after finishing questionnaire 2
18	Debriefing	3	Point B	Debriefing and thanked for participation	Both groups are told that the experiment may have caused negative emotions and stress. Control: they are asked to tell nothing about the experiment to other people. They can write their mail addresses down when they like to receive more information about the experiment. They will receive a mail later that day or at the next day about the purpose of the experiment. Experimental: they are debriefed about the purpose of the experiment. There

					is no further experiment conducted. Therefore, they can be fully informed about the objectives of the study.

Appendix D. Informed Consent

Informed consent, in line with the requirements of Fraunhofer and the requirements for projects supported by the European Union.

This privacy policy is about the experiment conducted by Fraunhofer IESE on November 17, 2015 within the scope of the EC Project 'Reliable and Smart Analysis of Crowdsourcing Information for Emergency and Crisis Management – EU' – RESCUER, Grant 614154.

Declaration of secrecy of your data

By participating in this study you support us in evaluating the RESCUER App. We affirm you that the data obtained with your help are solely used for the evaluation of the app and are treated in confidence. The data are obtained anonymized.

We certify, that a single questionnaire will never be given to a third-party nor will a third-party be able to link a test person to a data set. Third party means all persons, departments or companies that are not directly involved in this project.

We reserve to use the data in an anonymized manner for scientific papers, presentations, deliverables and other scientific publications. For this we would group and summarize the data.

Informed consent

This experiment may cause stress and/or distress. You have the right to withdraw your participation at every moment. Fraunhofer is not responsible for any potential damage or health interference that may arise during the experiment.

With your approval we will obtain data during this study. We obtain data related to certain tasks that you are asked to complete during this study. The data are, of course, obtained in an anonymized way. Obtaining data is done to gain knowledge for further improvements of the app and to check whether the way of evaluating the app was suitable. The obtained data include video recording of parts of the experiment as well as obtaining saliva samples.

O Yes, I agree that Fraunhofer IESE uses the obtained data for research and demonstration purposes in an anonymized way.

O Yes, I refrain from my right to check and control the data obtained within the project before the data are processed.

With this agreement you agree that your data are processed as described earlier in this document. Through the anonymous processing of your data we cannot draw any conclusions referring to your person. As a consequence of this, deleting your data afterwards is not possible.

Kaiserslautern, November 17, 2015

_____

_____

Name Signature

Declaration of commitment

I am informed that I have to promise secrecy about all affairs of the RESCUER project and Fraunhofer IESE, that contain for instance information about specifications, models, drawings, functionality of devices, objects or applications, even after the project has finished. This is also the case for all other affairs, I receive information about in the context of this project. I was explicitly informed about the legal requirements about unfair competition. I have to protect all notes, photographs, copies, I received or I made against the access of meddlers.

I am informed about these obligations. I am aware that it is a punishable offence to not fulfill those obligations, especially according to '§ 17 Gesetz gegen den unlauteren Wettbewerb (UWG)'. I know that in case of not fulfilling those obligations civil demands can be claimed from me.

Kaiserslautern, 17th of November 2015

_____ _____

Name Signature

Appendix E. Ethical Requirements

This documents provides information about the ethical constraints of studies dealing with stress. The ethical constraints are explained with regard to our study.

Ethical committee

Ethical guidelines for psychologists were created by national psychological associations (e.g. the American Psychological Association (APA) http://www.apa.org/ethics/code/index.aspx and the Deutsche Gesellschaft für Psychologie (DGP)) to ensure "welfare and protection of the individuals and groups with whom psychologists work" (http://www.apa.org/ethics/code/index.aspx). Local ethic committees approve or disapprove research proposals based on the national and local ethic codes, and the committee members own experience and opinion (for more information, check the SATORI project. It investigates among others the procedures of ethic committees in various countries). Our experiment has the objective to elicit stress. Stress is causing discomfort. To keep the discomfort at an acceptable level, we consult the research chapter of the ethical codes. The codes demand that participants have to sign an informed consent. Point 1 and 4 are of special interest for our experiment:

"Informed Consent, psychologists inform participants about (1) the purpose of the research, expected duration and procedures; (2) their right to decline to participate and to withdraw from the research once participation has begun; (3) the foreseeable consequences of declining or withdrawing; (4) reasonably foreseeable factors that may be expected to influence their willingness to participate such as potential risks, discomfort or adverse effects; (5) any prospective research benefits; (6) limits of confidentiality; (7) incentives for participation; and (8) whom to contact for questions about the research and research participants' rights." (http://www.apa.org/ethics/code/index.aspx)

Informing the participants about the purpose may change their behaviour, for instance, they may answer subjective questionnaires in a socially desirable way or in a way that the answers match the purpose of the study. Telling the participant that the experiment may create discomfort, maybe gives an indication what the study is about. Deviations of the informed consent are possible:

> "**8.05 Dispensing with Informed Consent for Research**
>
> Psychologists may dispense with informed consent only (1) where research would not reasonably be assumed to create distress or harm" (http://www.apa.org/ethics/code/index.aspx)

Since the experiment will probably create distress, the participants have to sign an informed consent. Still, there are deviations in telling the purpose of the study.

“8.07 Deception in Research

(a) Psychologists do not conduct a study involving deception unless they have determined that the use of deceptive techniques is justified by the study's significant prospective scientific, educational or applied value and that effective nondeceptive alternative procedures are not feasible. (b) Psychologists do not deceive prospective participants about research that is reasonably expected to cause physical pain or severe emotional distress.” (http://www.apa.org/ethics/code/index.aspx)

Whether the experiment causes severe emotional distress or an acceptable level of stress, is a question for an ethic committee. The stress tests TSST, MAST, SSST, SECPT, and MMST all mention that the test was approved by a local ethic committee. The study about MMST (Kolotylova et al., 2008) mentions that stress-related words were avoided in the informed consent. The other stress tests do not mention any adaptations of the informed consent.

The duration of the experiment is maybe a reason to not approve the experiment. In our study, the stress manipulation lasts for about 5 minutes. The duration of the SECPT, which makes use of thermal pain, lasted for a maximum of three minutes, MAST induces stress for five minutes, and TSST even for 20 minutes. Thus, the duration of the experiment does not exceed other approved stress protocols.

After the experiment, psychologists are still responsible for the well-being of their participants.

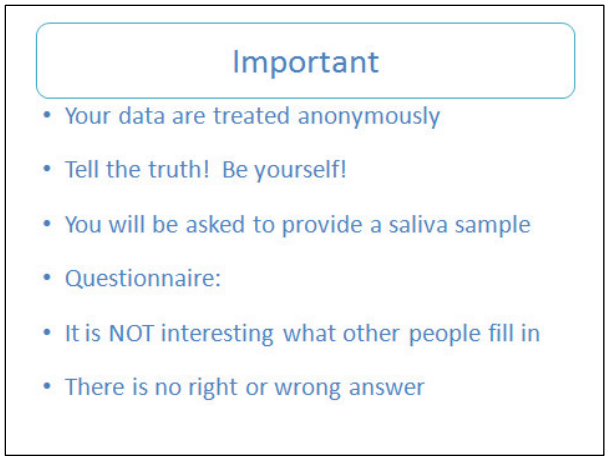
“8.08 Debriefing

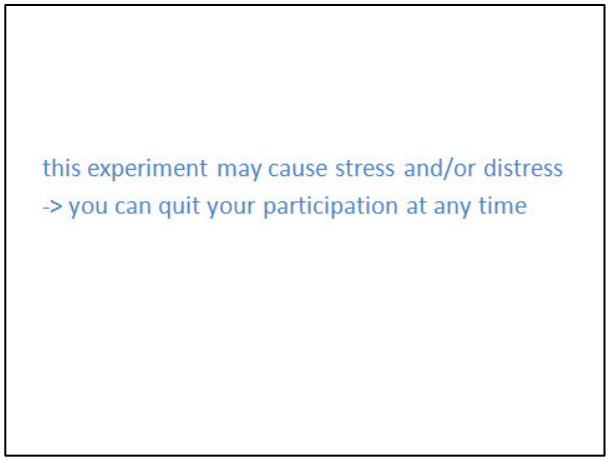
(…) (c) When psychologists become aware that research procedures have harmed a participant, they take reasonable steps to minimize the harm.”

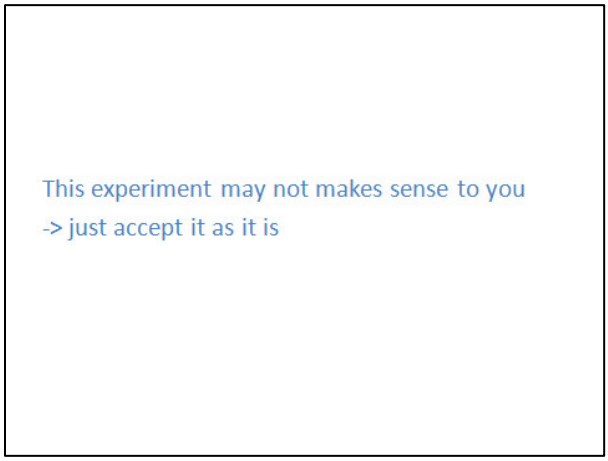
The studies about the development and testing of TSST, MAST, and MMST do not take special procedures to reduce stress after the experiment (Kirschbaum et al., 1993; Reinhardt et al., 2012; Smeets et al., 2012). The participants are instructed to relax or may do non-stressful tasks until the final cortisol measure (after 40 minutes). We do not think that the experiment will harm the participants. Still, the participants receive some more information about the experiment with an explanation that the experienced emotions and stress are related to the experiment.

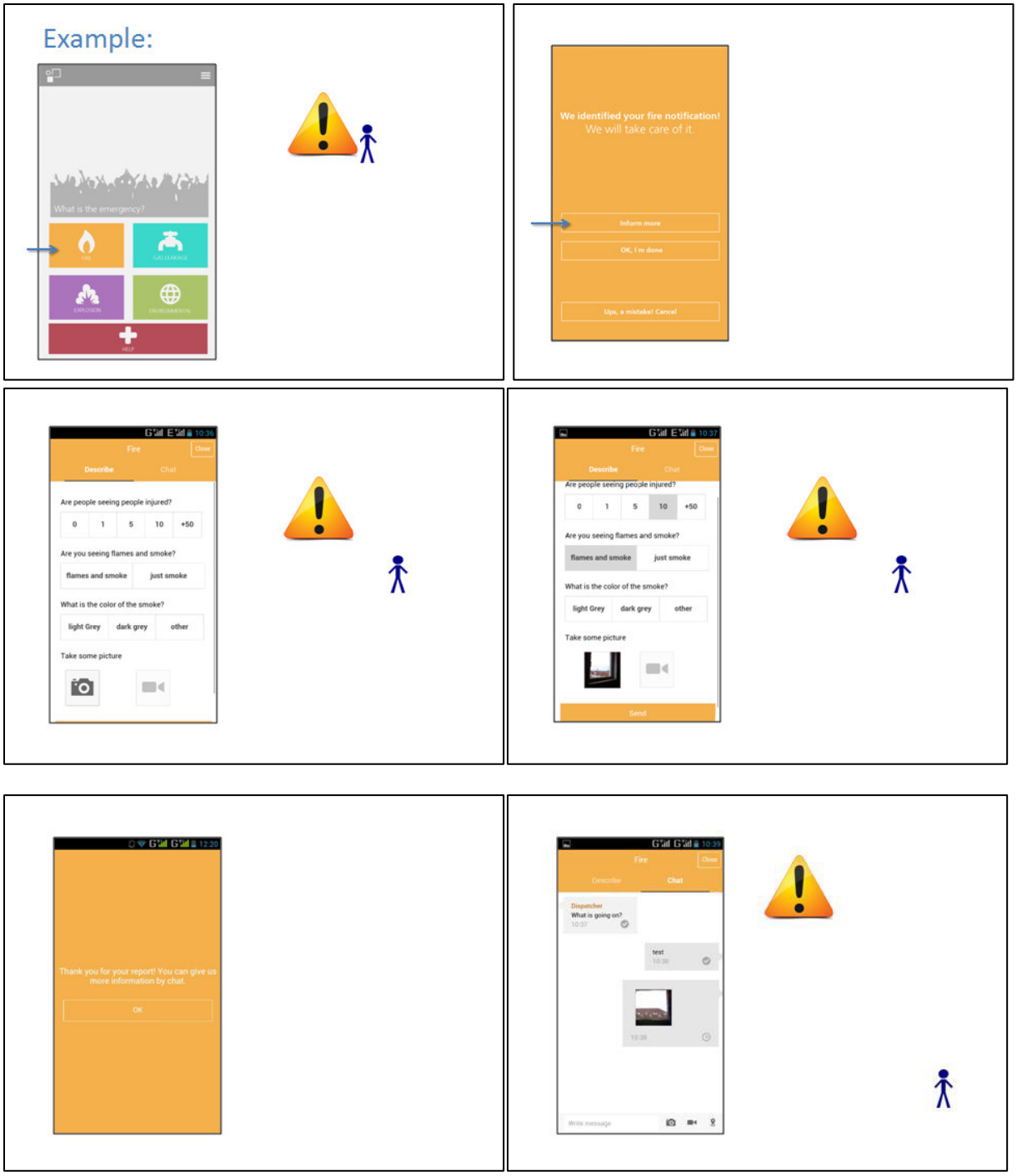
Appendix F. Presentation about the Experiment

The slides show parts of the presentation that was given before the experiment to explain some basics about psychological experiments and to introduce the app. It was explained to the participants that the triangle represents an incident and the figure the user. The distance between user and incident was meant to give an indication when to use the different screens.








Example:
What is the emergency?
We identified your fire notification!
We will take care of it.
Inform more
OK, I'm done
Ups, a mistake! Cancel
Fire
Close
Describe
Chat
Are people seeing people injured?
0
1
5
10
+50
Are you seeing flames and smoke?
flames and smoke
just smoke
What is the color of the smoke?
light Grey
dark grey
other
Take some picture
Send
Thank you for your report! You can give us more information by chat.
OK
Dispatcher
What is going on?
test
Write message

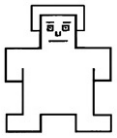
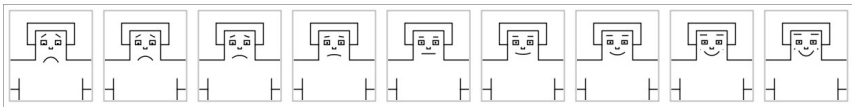
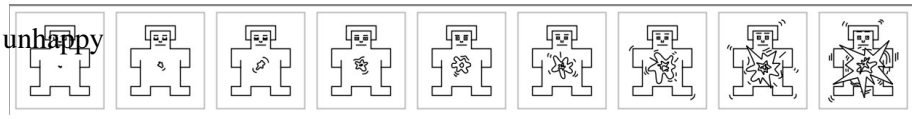
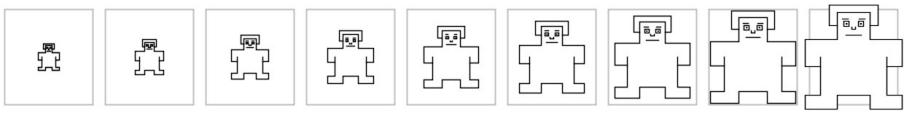
Appendix G. Questionnaires

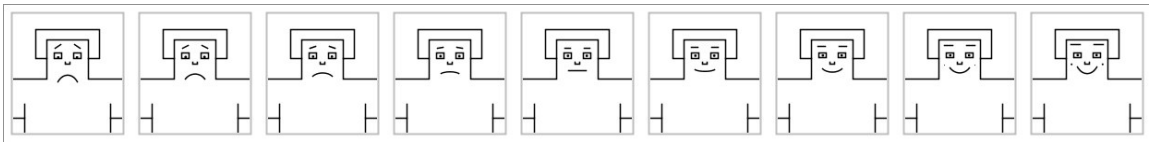
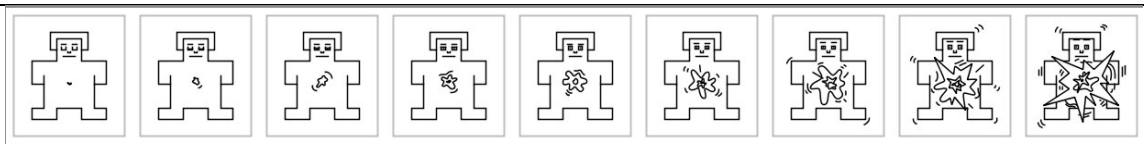
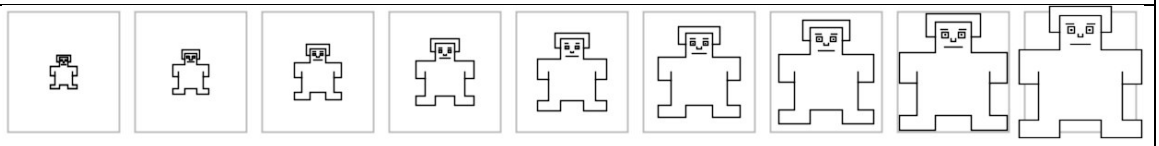
The first questionnaire was filled in before the experiment started and the second one after the experiment finished. The introduction of the SAM was taken from the former evaluation of RESCUER App (Nass et al., 2015).

Questionnaire 1

Please, fill in the questionnaire at your own pace. It is very important that you answer the questions *honestly*. Your data will be treated anonymously. There are *no* right or wrong answers, *neither* is the questionnaire used to evaluate you personally.

Demographic data	
1.	What is your participant's number? _____
2	I am: O male O female
3.	What is your age? _____
4.	What is your nationality? _____
	My understanding of the purpose of this experiment is: O _____ O I don't know
5.	What is your highest educational attainment: O Doctorate O University degree (Bachelor, Master) O Higher education entrance qualification (Abitur etc.) O Secondary school leaving certificate (MittlereReife) O Other:_____
5a.	I am a(n): O employee at Fraunhofer O student working for Fraunhofer or writing a thesis there O student (not related to Fraunhofer) O others: _____
5b.	If you are related to Fraunhofer: Are you involved in the RESCUER project? O Yes

	O No, but I know what the project is about O No, and I do *not* know what the project is about
5c.	Do you know the RESCUER App? O Yes, and I have used/tested it before O Yes, but I have only seen it on screenshots or as a demo O No, I do not know it
6.	What kind of smartphone operating system do you use regularly? (several answers possible) O Android O iOS (iPhone) O Windows Phone
Your current state	
7.	How stressed do you feel at the moment? No stress O 0 O 1 O 2 O 3 O 4 O 5 O 6 O 7 O 8 O 9 O 10 Extremely stressed
8.	Introduction to SAM. Only read this text. You will be asked later on at question 8a to 8c to mark the pictures. This is SAM:  You will see him several times in this experiment. You will use SAM to express your mood. The first scales 'pleasure' ranges from a frowning, unhappy face to a happy face with a big smile.  unhappy happy The second scale measures your level of arousal ranges from a SAM with sleepy face and closed eyes to a very excited SAM.  calm Very excited The third scale is about dominance. The left picture depicts a feeling of being *controlled* /having no control over the situation. The right picture illustrates a feeling of being *in control* of the situation.  no control In control
8a.	We are interested in how you feel **right now**. There is no right or wrong answer. Please, answer

	spontaneously by marking the pictures with a cross or by making a circle around them. How do you feel right now?  O O O O O O O O O
8b.	 O O O O O O O O O
8c.	 O O O O O O O O O

Some more questions

9a.	Do you have any experience with emergencies? An emergency is a situation that required calling the emergency services (e.g., police, ambulance, fire department) O No, I do *not* have experience with emergencies O Yes, I have experience with emergencies.
9b.	The emergency was a: O car accident O fire O gas leak O explosion O others: _____
9c.	I called the emergency services ___ time(s)
10.	Which of the following activities did you perform in the last hour? O Intensive physical exercise (e.g., jogging) O Drinking a beverage with caffeine (e.g., coffee, black tea, energy drink) O Consuming drinks and/or foodstuffs O Brushing teeth O Smoked a cigarette O I performed none of those activities
11.	Have you taken any medications that may still be active in your body? O No O Yes, namely: _____
12.	Do you have any medical conditions that you expect can affect your performance (using an

	app and walking) in this experiment? □ No □ Yes, and it can affect my performance in this experiment in the following way(s): _____

Questionnaires 2

The second questionnaire was filled in after the experiment.

The questionnaires are presented in the following order:

1. Questionnaire for the control group with the app (p.)
2. Questionnaire for the control group without the app (p.)
3. Questionnaire for the experimental group with the app (p.)
4. Questionnaire for the experimental group without the app (p.)
   The questionnaire of this group contains one wrong screenshot of an interaction screen.

What is your participant's number?

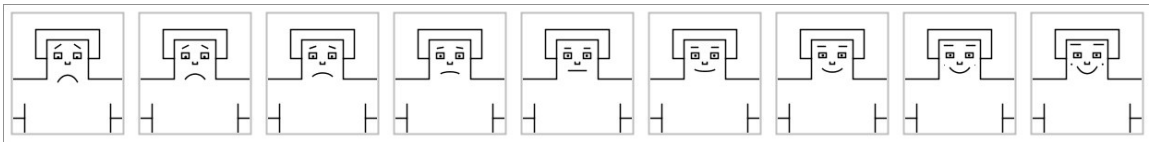
_____

Current state

1. How stressed did you feel **during the experiment**?
During =when going from the starting room to point B (where you are at this moment)

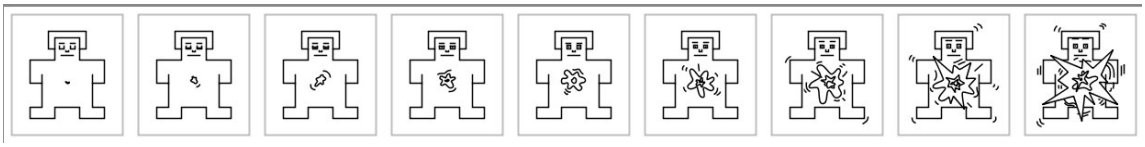
No stress											Extremely stressed
O	O	O	O	O	O	O	O	O	O	O	
0	1	2	3	4	5	6	7	8	9	10	

2a. How did you feel **during the experiment?**



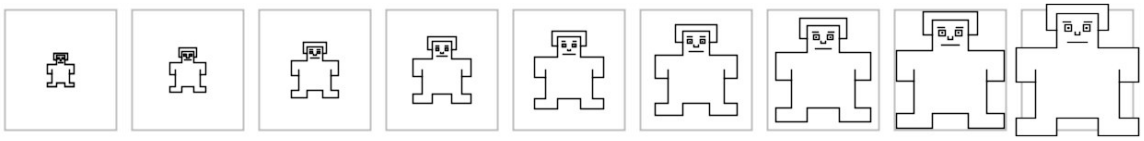
O O O O O O O O O

2b. How did you feel **during the experiment?**



O O O O O O O O O

2c. How did you feel **during the experiment?**

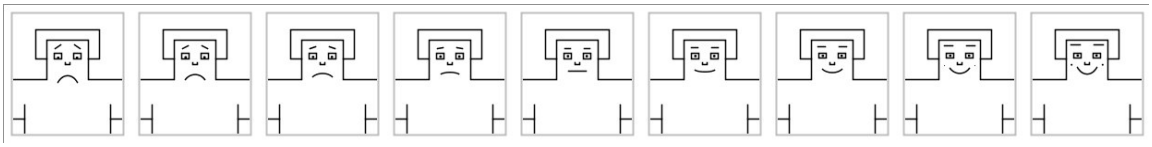


O O O O O O O O O

3. How stressed do you feel **at this moment**:

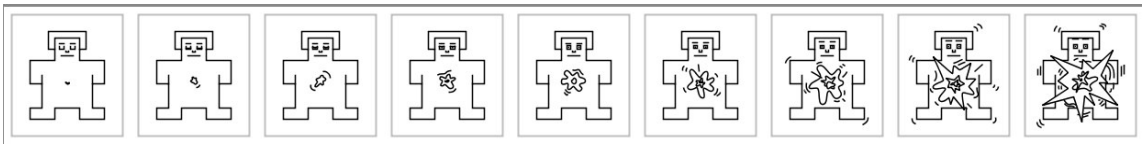
No stress											Extremely stressed
O	O	O	O	O	O	O	O	O	O	O	
0	1	2	3	4	5	6	7	8	9	10	

4a. How do you feel **at this moment**?



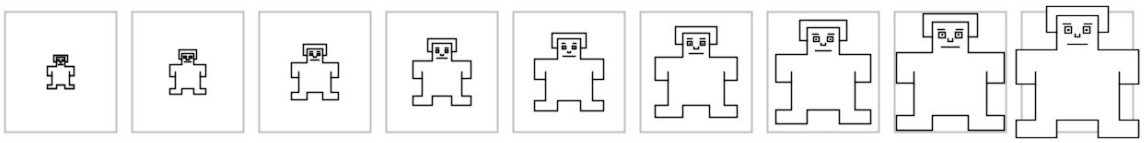
O O O O O O O O O

4b. How do you feel **at this moment**?



O O O O O O O O O

4c. How do you feel **at this moment**?



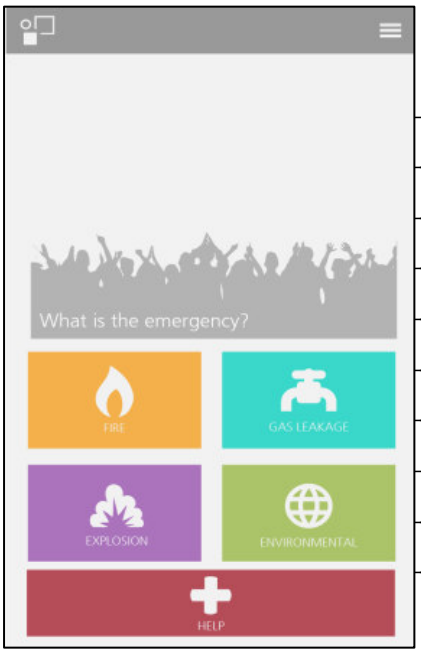
O O O O O O O O O

Questions about the RESCUER App

The goal of this evaluation is to detect weaknesses and strengths of the RESCUER App and to develop concrete suggestions for improvement.

- In order to achieve this, your judgement as (potential) user of the RESCUER App program is of paramount importance. Your evaluation of the RESCUER App should be based on your personal experience.
- It is important to remember that this is not an evaluation of you personally. Rather, we are interested in your personal evaluation of RESCUER App
- When you do not have any comment to the open questions, please, cross the lines for the answer, so that I know that you did not answer the question on purpose.

5.



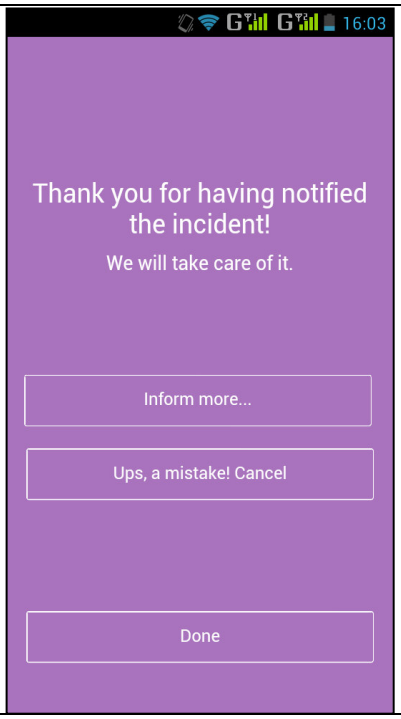
a) What do you like about this screen?

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

b) What do you not like about this screen/ What should be improved?

_____
_____
_____
_____
_____
_____
_____

<table>
<tr><td></td><td>c) Do you have any other comments about this screen?

_____
_____
_____
_____
_____
_____
_____</td></tr>
<tr><td>6</td><td> 

a) What do you like about this screen?

_____
_____
_____
_____

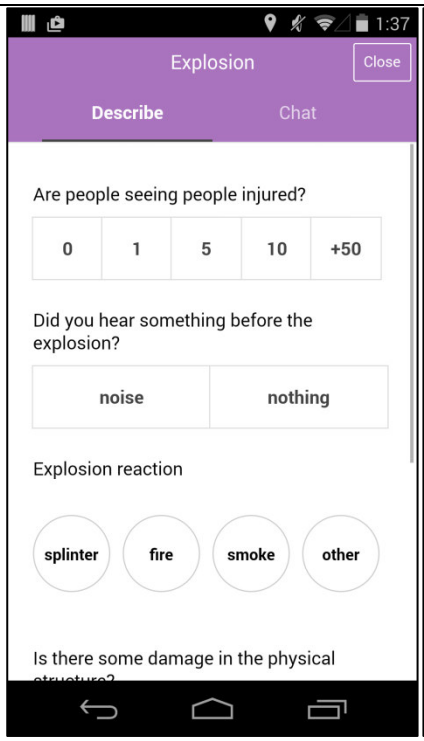
b) What do you not like about this screen? What should be improved?

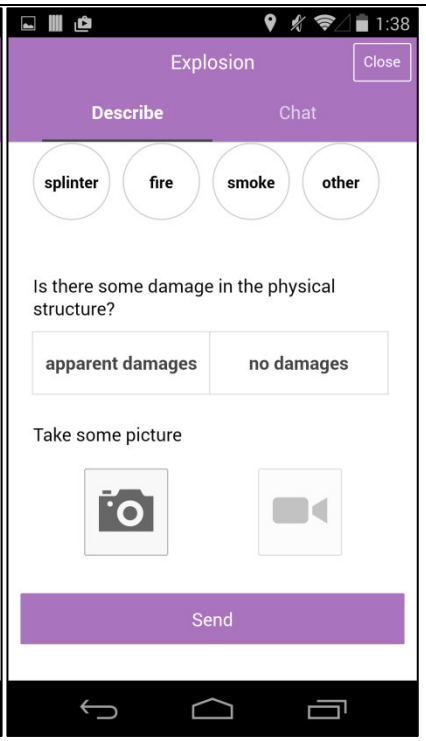
_____
_____
_____
_____

c) Do you have any other comments?

_____
_____</td></tr>
</table>

7.





Upper part of the screen lower part (when scrolling down)

a) What do you like about this screen?

_____

_____

_____

_____

_____

_____

a) What do you not like about this screen/ What should be improved?

_____

_____

_____

_____

_____

_____

_____

b) The questions at the screen are

clear O O O O O unclear

continue on the next page

c) Do you have any other comments about the questions in the app?

_____

_____

_____

_____

d) It is clear that I have to click on the send button to send my data:
O yes O no

e) The answer possibilities are in the app

clear O O O O O unclear

f) Do you have any other comments about the answer possibilities?

_____

_____

_____

_____

_____

g) Do you have any other comments about this screen?

_____

_____

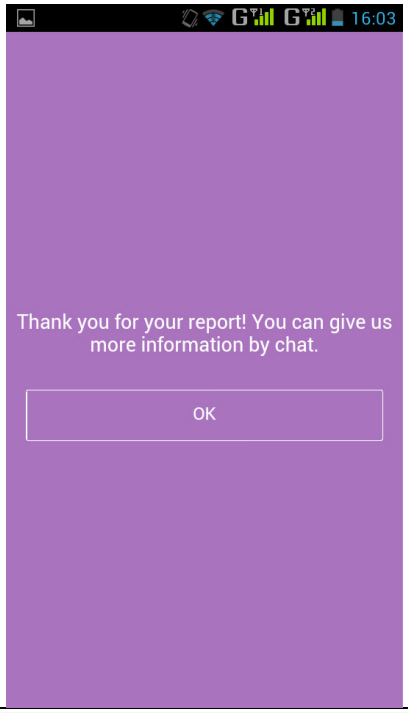
_____

_____

_____

Continue on the next page

8



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

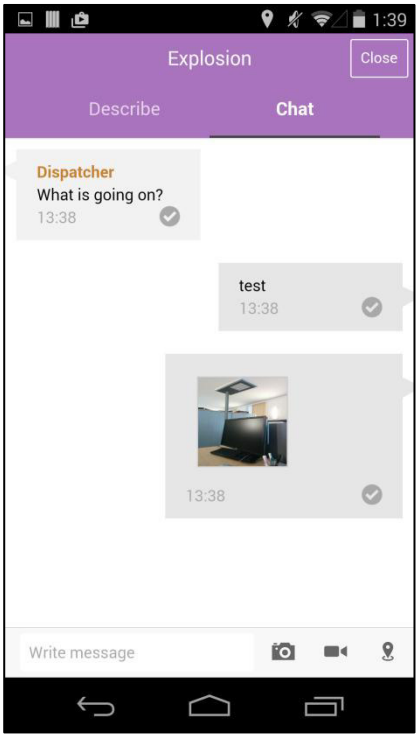
_____

c) Do you have any other comments?

_____

_____

9.



a) What do you like about this screen of the chat function?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

b) What do you not like about this screen?/ What should be improved?

_____

_____

_____

_____

_____

_____

_____

c) Do you have any other comments about this screen?

_____

_____

_____

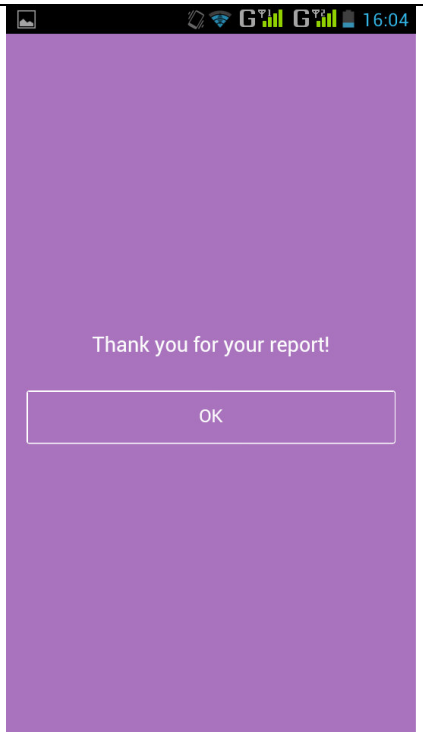
_____

_____

_____

_____

10.



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

_____

c) Do you have any other comments?

_____

_____

The RESCUER App…	---	--	-	-/+	+	++	+++	**The RESCUER App…**
Requires unnecessary input.	O	O	O	O	O	O	O	Requires unnecessary input.
provides insufficient information about which entries are valid and necessary.	O	O	O	O	O	O	O	provides sufficient information about which entries are valid and necessary.
complicates orientation due to an inconsistent design	O	O	O	O	O	O	O	facilitates orientation due to a consistent design.

The RESCUER App…	--- -- - -/+ + ++ +++	**The RESCUER App…**
requires the memorization of too many details.	O O O O O O O	does not require the memorization of too many details.
requires a lot of time to learn	O O O O O O O	requires little time to learn.
is difficult to learn without external support or a handbook	O O O O O O O	is easy to learn without external support or a handbook.

11.	Would you use the app in a real emergency? O yes, because _____ O no, because_____ O I don't know, because ___
12.	What do you think is the purpose of the experiment? _____
13.	Do you have any other comment about the experiment or the RESCUER App? _____ _____ _____ _____ _____

Thank you very much for participating in this study!

What is your participant number?

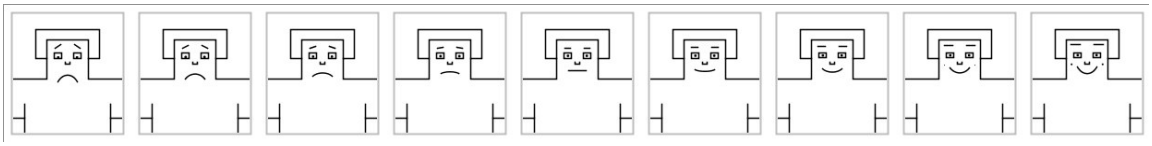
_____

Current state

1. How stressed did you feel **during the experiment**?
During =when going from the starting room to point B (where you are at this moment)

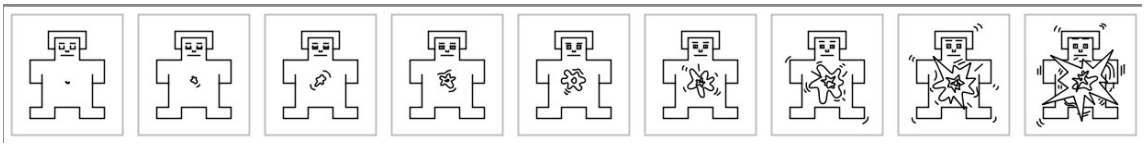
No stress											Extremely stressed
O	O	O	O	O	O	O	O	O	O	O	
0	1	2	3	4	5	6	7	8	9	10	

2a. How did you feel **during the experiment?**



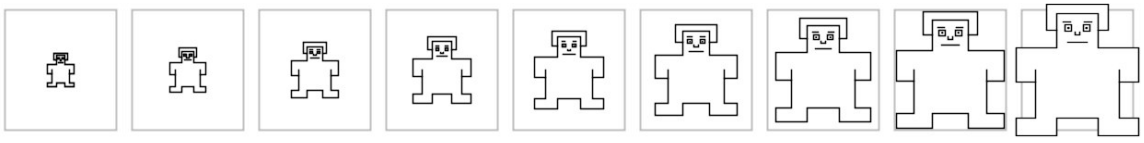
O O O O O O O O O

2b. How did you feel **during the experiment?**



O O O O O O O O O

2c. How did you feel **during the experiment?**

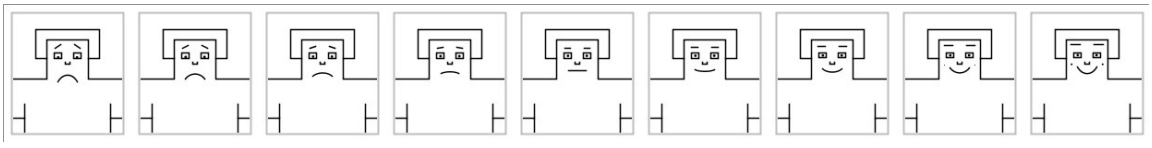


O O O O O O O O O

3. How stressed do you feel **at this moment**?

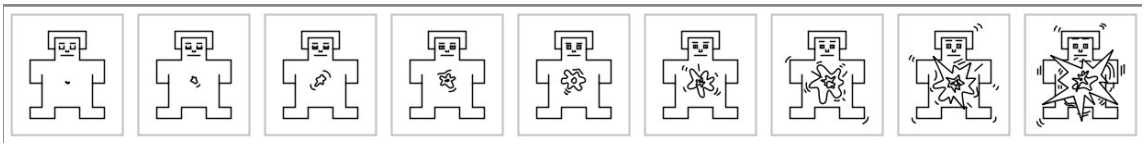
No stress											Extremely stressed
O	O	O	O	O	O	O	O	O	O	O	
0	1	2	3	4	5	6	7	8	9	10	

4a. How do you feel **at this moment**?

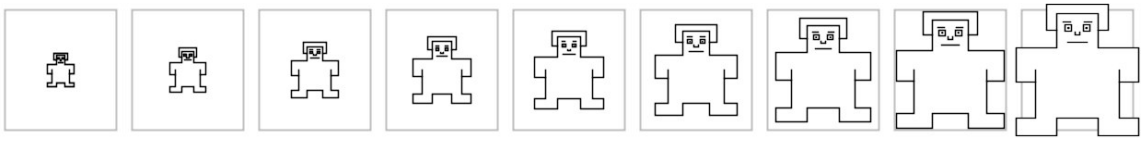


O O O O O O O O O

4b. How do you feel **at this moment**?



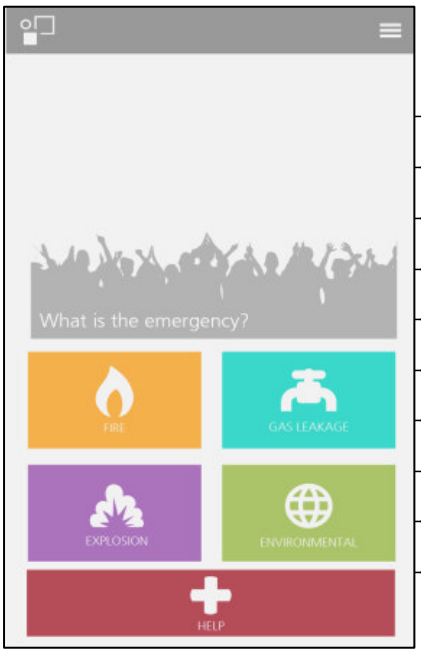
O O O O O O O O O

4c.	How do you feel **at this moment**?  O O O O O O O O O
Questions about the incident	
5.	Now that you left the hazard zone the rescue workers need more information about what has happened. They prepared these questions for you: What is the emergency? O fire O gas O explosion O environmental Are people seeing people injured?(Circle the answer) 0 1 5 10 +50 Did you hear something before the explosion? O noise O nothing Explosion reaction O splinter O fire O smoke O other Is there some damage in the physical structure? O apparent damage O no damages Write about what was going on if you would like to give more information: _____ _____ _____ _____ _____ _____
Questions about the RESCUER App	
The goal of this evaluation is to detect weaknesses and strengths of the RESCUER App and to develop concrete suggestions for improvement. • In order to achieve this, your judgement as (potential) user of the RESCUER App program is of paramount importance. Your evaluation of the RESCUER App should be based on your personal experience.	

• It is important to remember that this is not an evaluation of you personally. Rather, we are interested in your personal evaluation of the RESCUER App

• When you do not have any comment to an open question, please, cross the lines for the answer, so that I know that you did not answer on purpose.

6.

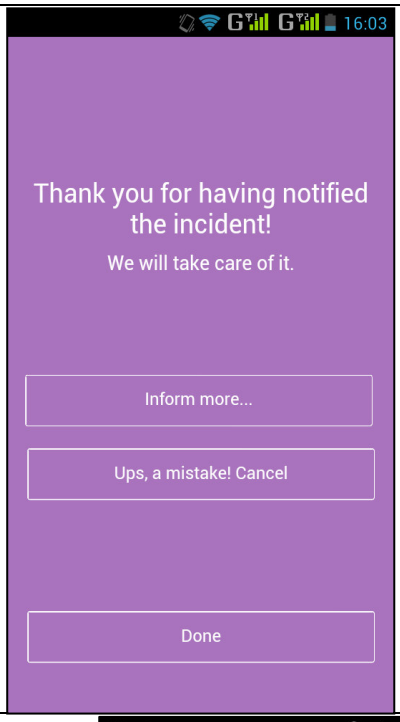


a) What do you like about this screen?

b) What do you not like about this screen/ What should be improved?

d) Do you have any other comments about this screen?

7.



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

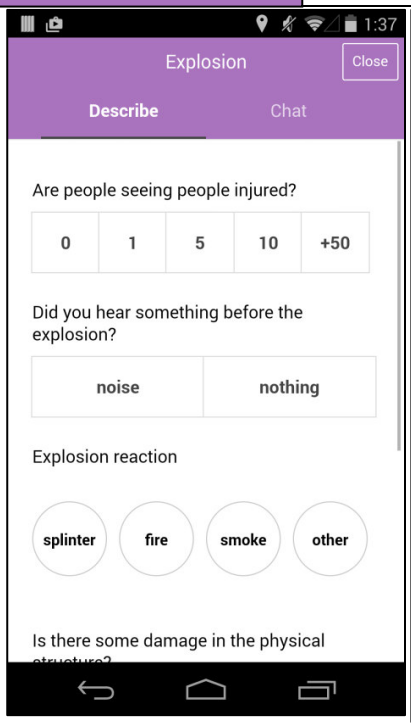
_____

c) Do you have any other comments?

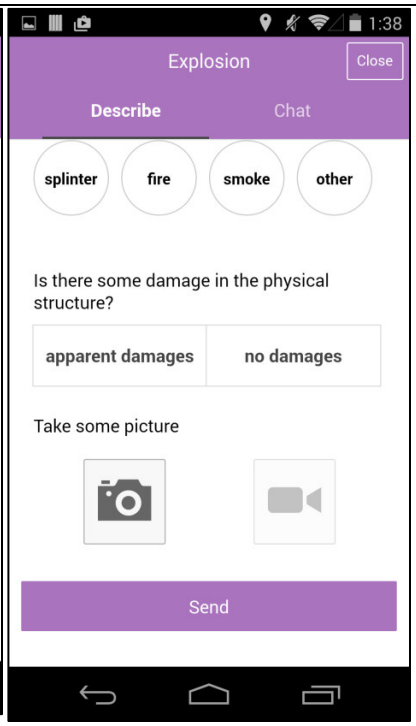
_____

_____

8.



Upper part of the screen



lower part (when scrolling down)

d) What do you like about this screen?

_____

_____

_____

_____

_____

_____

e) What do you not like about this screen/ What should be improved?

_____

_____

_____

_____

_____

_____

_____

f) The questions at the screen are

clear O O O O O unclear

g) Do you have any other comments about the questions?

_____

_____

_____

_____

h) It is clear that I have to click on the send button to send my data:
O yes O no

i) The answer possibilities are in the app

clear O O O O O unclear

j) Do you have any other comments about the answer possibilities?

_____

_____

_____

_____

_____

k) Do you have any other comments about this screen?

_____

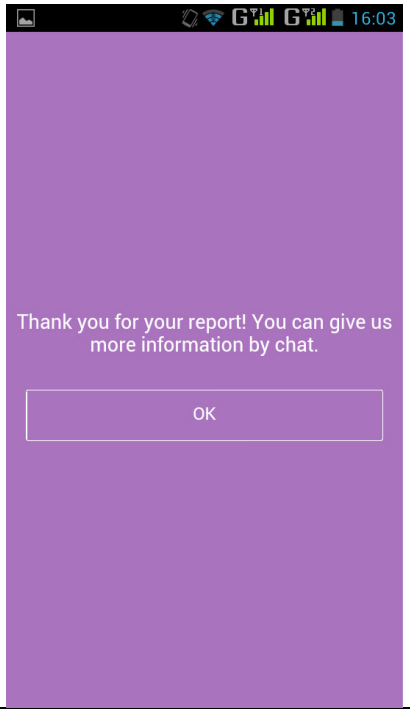
_____

_____

_____

_____

9.

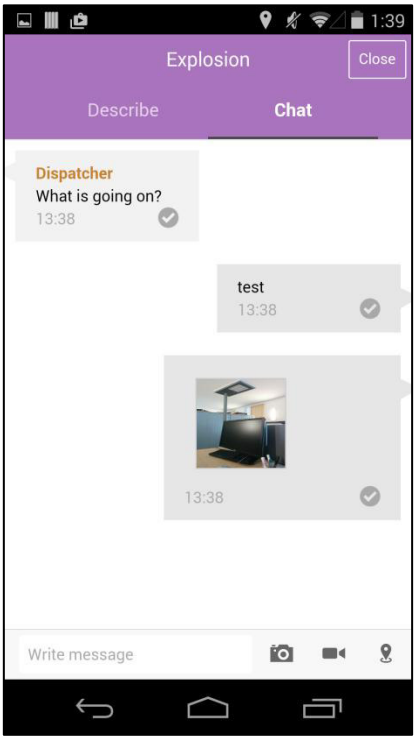


a) What do you like about this screen?

b) What do you not like about this screen? What should be improved?

c) Do you have any other comments?

10.



a) What do you like about the screen of the chat function?

b) What do you not like about this screen/ What should be improved?

c) Do you have any other comments about this screen?

_____

_____

_____

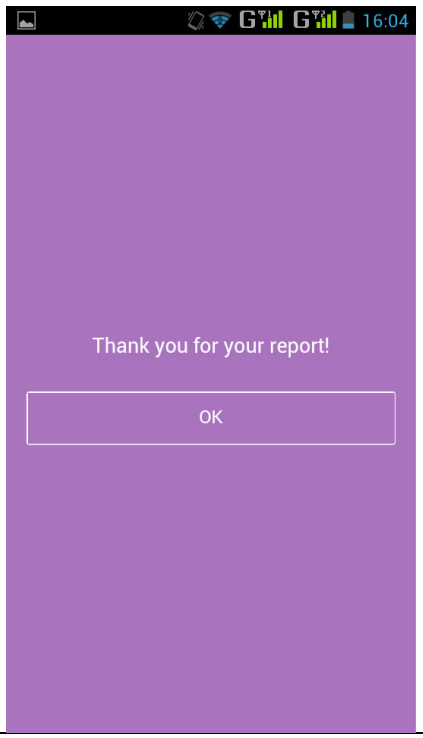
_____

_____

_____

_____

11.



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

_____

c) Do you have any other comments?

_____

_____

The RESCUER App…	--- -- - -/+ + ++ +++	**The RESCUER App…**
Requires unnecessary input.	O O O O O O O	does not require unnecessary input.
provides insufficient information about which entries are valid and necessary.	O O O O O O O	provides sufficient information about which entries are valid and necessary.
complicates orientation due to an inconsistent design.	O O O O O O O	facilitates orientation due to a consistent design.

The RESCUER App…	--- -- - -/+ + ++ +++	**The RESCUER App…**
requires the memorization of too many details.	O O O O O O O	does not require the memorization of too many details.
requires a lot of time to learn	O O O O O O O	requires little time to learn.
is difficult to learn without external support or a handbook	O O O O O O O	is easy to learn without external support or a handbook.

12.	Would you use the app in a real emergency? O yes, because _____ O no, because_____ O I don't know, because ___ _____
13	What do you think is the purpose of the experiment? _____
14.	Do you have any other comments about the experiment or the RESCUER App? _____ _____ _____ _____ _____

What is your participant number?

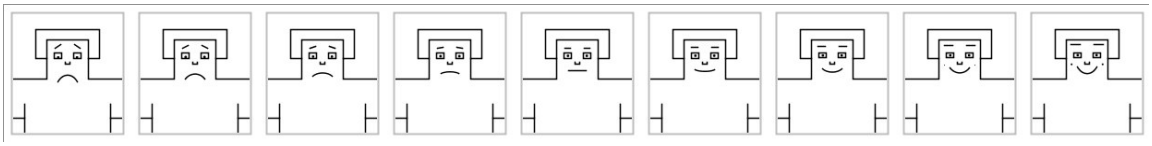
_____

Current state

1. How stressed did you feel **during the experiment** ?
During =when going from the starting room to point B (where you are at this moment)

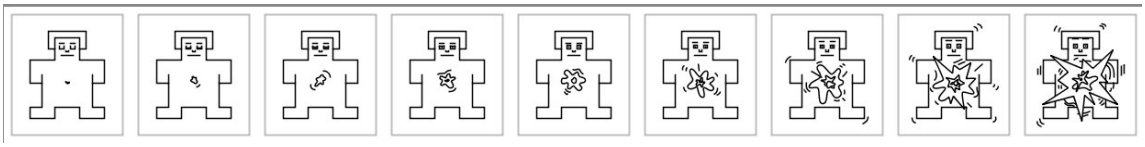
No stress											Extremely stressed
O	O	O	O	O	O	O	O	O	O	O	
0	1	2	3	4	5	6	7	8	9	10	

2a. How did you feel **during the experiment?**



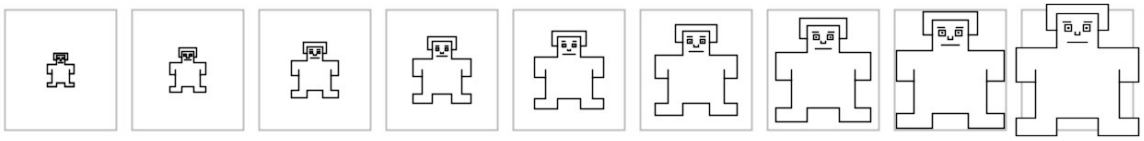
O O O O O O O O O

2b. How did you feel **during the experiment?**



O O O O O O O O O

2c. How did you feel **during the experiment?**

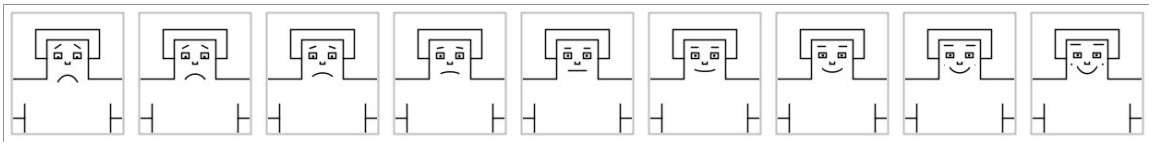


O O O O O O O O O

3. How stressed do you feel **at this moment**:

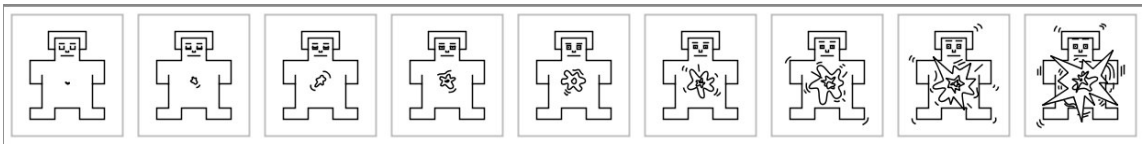
No stress											Extremely stressed
O	O	O	O	O	O	O	O	O	O	O	
0	1	2	3	4	5	6	7	8	9	10	

4a. How do you feel **at this moment**?



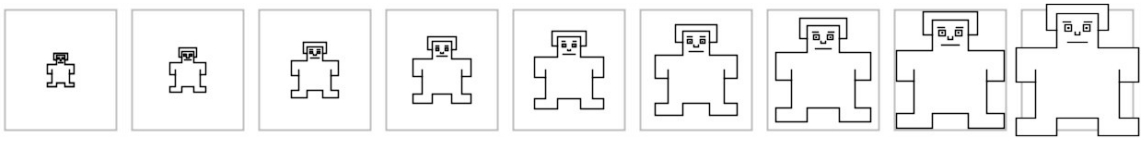
O O O O O O O O O

4b. How do you feel **at this moment**?



O O O O O O O O O

4c. How do you feel **at this moment**?



O O O O O O O O O

Question about the experiment

5. As how stressful did you experience the following elements of the experiment?

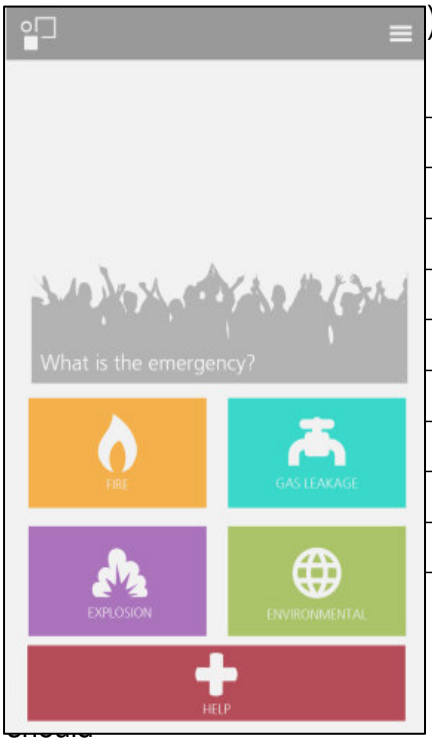
	not stressful	low	medium	high	very stressful
Sound of screaming people	O	O	O	O	O
Sound of fire alarm	O	O	O	O	O
Presentation about an explosion	O	O	O	O	O
Watching the video of the explosion	O	O	O	O	O
Pictures of humans	O	O	O	O	O
Pictures of explosions	O	O	O	O	O
Persons telling the way	O	O	O	O	O
Other persons in the group	O	O	O	O	O
Walking to point B	O	O	O	O	O
Finding the way to point B	O	O	O	O	O
Using the app	O	O	O	O	O

Questions about the RESCUER App

The goal of this evaluation is to detect weaknesses and strengths of the RESCUER App and to develop concrete suggestions for improvement.

- In order to achieve this, your judgement as (potential) user of the RESCUER App is of paramount importance. Your evaluation of the RESCUER App should be based on your personal experience.
- It is important to remember that this is not an evaluation of you personally. Rather, we are interested in your personal evaluation of RESCUER App
- When you do *not* have any comment to the open questions, please, cross the lines for the answer, so that I know that you did not answer the question on purpose.

6.

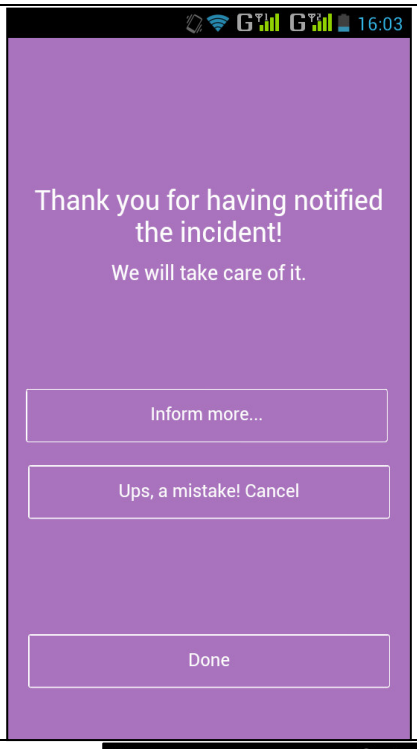


) a) What do you like about this screen?

b) What do you not like about this screen/ What should be improved?

e) Do you have any other comments about this screen?

Continue on the next page



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

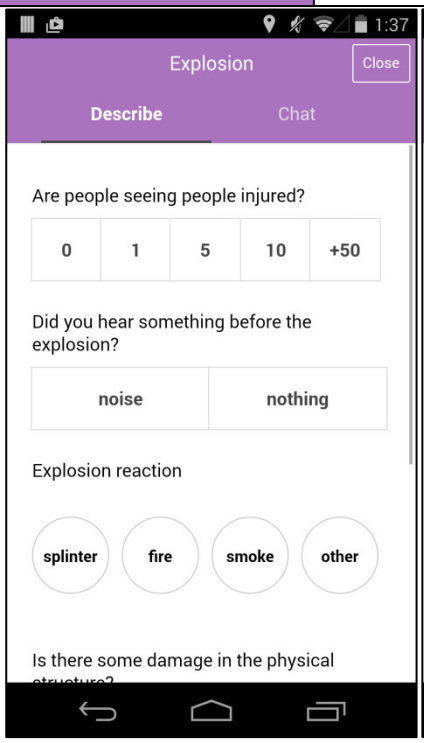
_____

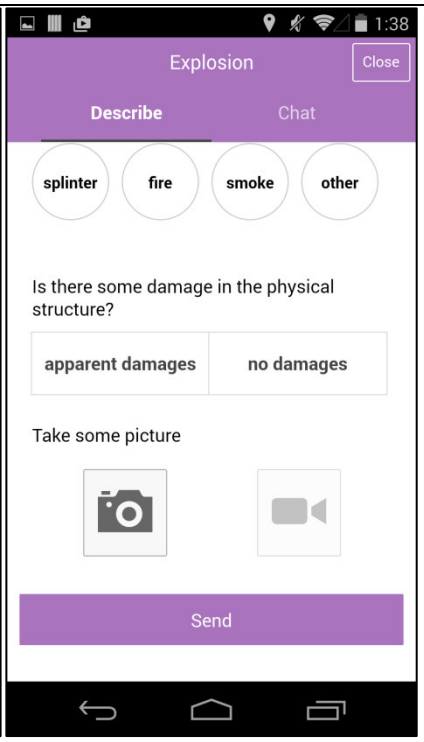
c) Do you have any other comments?

_____

_____

7.





Upper part of the screen lower part (when scrolling down)

I) What do you like about this screen?

_____

_____

_____

_____

_____

_____

m) What do you not like about this screen/ What should be improved?

_____

_____

_____

_____

_____

_____

_____

n) The questions at the screen are

clear O O O O O unclear

o) Do you have any other comments about the questions in the app?

_____

_____

_____

_____

p) It is clear that I have to click on the send button to send my data:
O yes O no

q) The answer possibilities are in the app

clear O O O O O unclear

r) Do you have any other comments about the answer possibilities?

_____

_____

_____

_____

_____

s) Do you have any other comments about this screen?

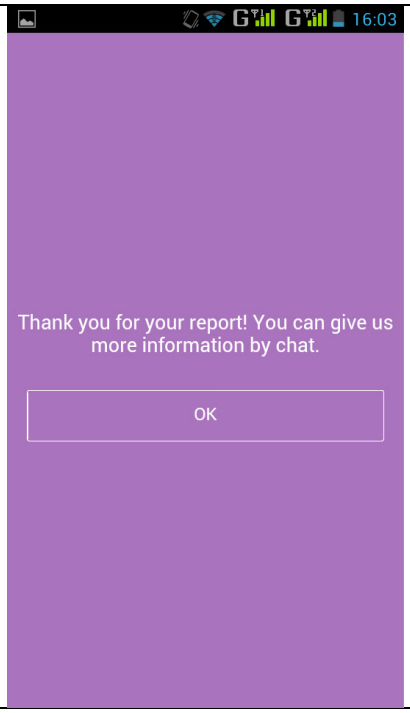
_____

_____

_____

_____

_____



a) What do you like about his screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

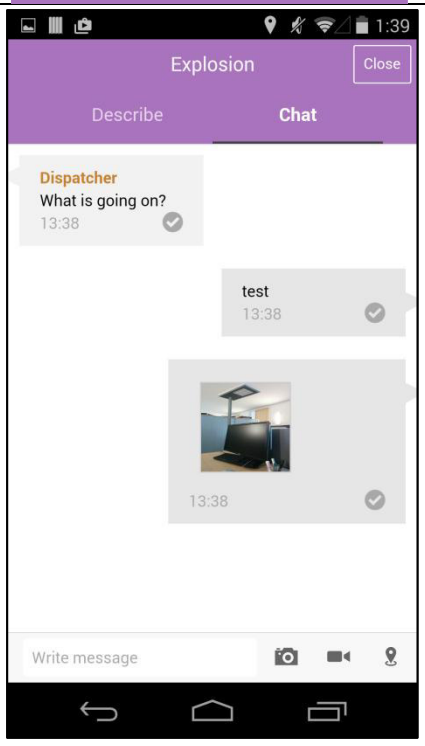
_____

c) Do you have any other comments?

_____

_____

8.



a) What do you like about the screen of the chat function?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

b) What do you not like about this screen?/ What should be improved?

_____

_____

_____

_____

_____

_____

_____

c) Do you have any other comments about the chat function?

_____

_____

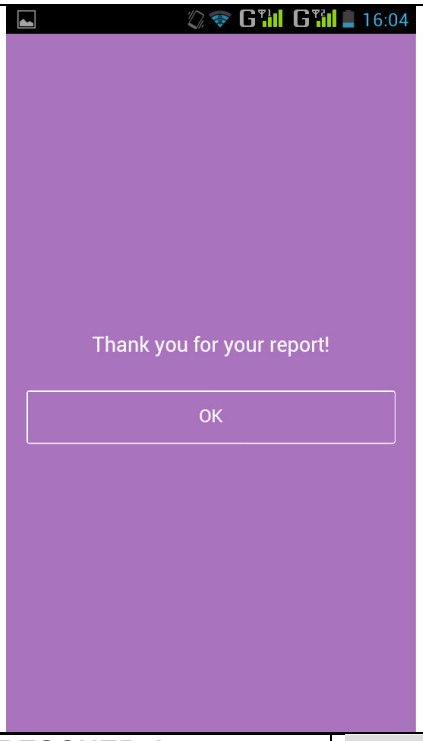
_____

_____

_____

_____

_____



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

_____

c) Do you have any other comments?

_____

_____

The RESCUER App…	--- -- - -/+ + ++ +++	**The RESCUER App…**
Requires unnecessary input.	O O O O O O O	does not require unnecessary input.
provides insufficient information about which entries are valid and necessary.	O O O O O O O	provides sufficient information about which entries are valid and necessary.
complicates orientation due to an inconsistent design	O O O O O O O	facilitates orientation due to a consistent design.

The RESCUER App…	--- -- - -/+ + ++ +++	**The RESCUER App…**
requires the memorization of too many details.	O O O O O O O	does not require the memorization of too many details.
requires a lot of time to learn	O O O O O O O	requires little time to learn.
is difficult to learn without external support or a handbook	O O O O O O O	is easy to learn without external support or a handbook.
provides error messages whichare difficult to understand.	O O O O O O O	provides error messages which are easy to understand.
does not give concrete help for error correction.	O O O O O O O	gives concrete support for error correction.

10	Would you use the app in a real emergency? O yes, because _____ O no, because_____ O I don't know, because ___ _____
11.	What do you think is the purpose of the experiment? _____
12.	Do you have any other comments about the experiment or the app? _____ _____ _____ _____ _____

What is your participant number?

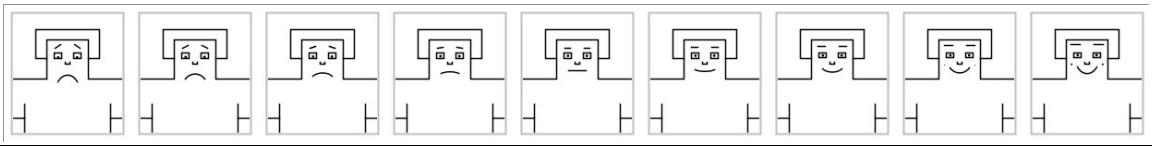
_____

Current state

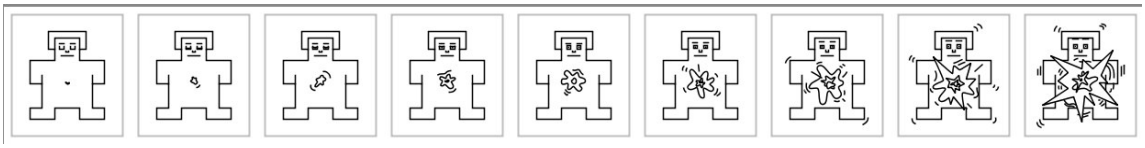
1. How stressed did you feel **during the experiment** ?
During =when going from the starting room to point B (where you are at this moment)

No stress	O 0	O 1	O 2	O 3	O 4	O 5	O 6	O 7	O 8	O 9	O 10	Extremely stressed

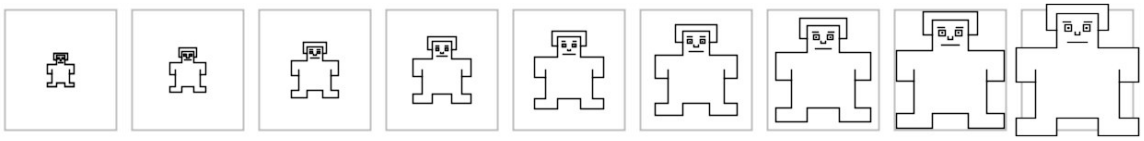
2a. How did you feel **during the experiment?**



2b. How did you feel **during the experiment?**



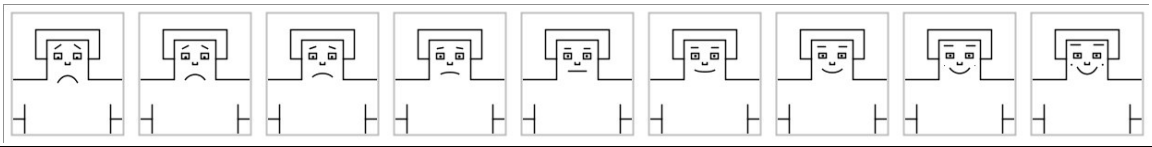
2c. How did you feel **during the experiment?**



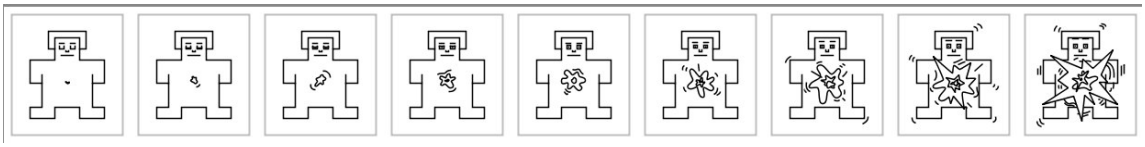
3. How stressed do you feel **at this moment**:

No stress	O 0	O 1	O 2	O 3	O 4	O 5	O 6	O 7	O 8	O 9	O 10	Extremely stressed

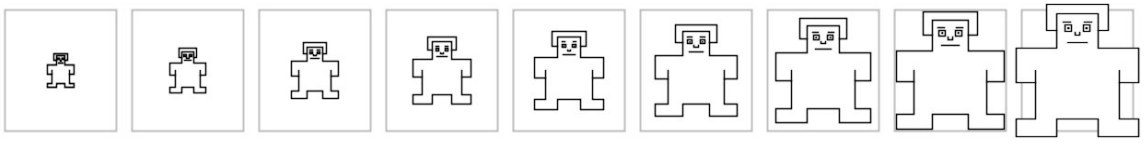
4a. How do you feel **at this moment**?



4b. How do you feel **at this moment**?



4c. How do you feel **at this moment**?



Questions about the incident

5. Now that you left the hazard zone the rescue workers need more information about what has happened. They prepared these questions for you:

What is the emergency?
O fire
O gas
O explosion
O environmental

Are people seeing people injured?(Circle the answer)
0 1 5 10 +50

Did you hear something before the explosion?
O noise O nothing

Explosion reaction
O splinter O fire O smoke O other

Is there some damage in the physical structure?
O apparent damage O no damages

Write about what was going on if you would like to give more information:

_____

_____

_____

_____

_____

_____

Questions about the experiment

6.

	not stressful	low	medium	high	very stressful
Sound of screaming people	O	O	O	O	O
Sound of fire alarm	O	O	O	O	O
Presentation about an explosion	O	O	O	O	O
Watching the video of the explosion	O	O	O	O	O
Pictures of humans	O	O	O	O	O
Pictures of explosions	O	O	O	O	O

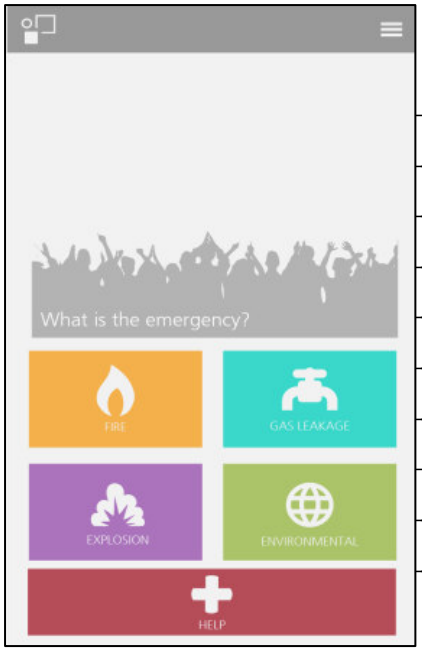
	not stressful	low	medium	high	very stressful
Persons telling the way	O	O	O	O	O
Other persons in the group	O	O	O	O	O
Walking to point B	O	O	O	O	O
Finding the way to point B	O	O	O	O	O

Questions about the RESCUER App

The goal of this evaluation is to detect weaknesses and strengths of the RESCUER App and to develop concrete suggestions for improvement.

- In order to achieve this, your judgement as a (potential) user of the RESCUER App program is of paramount importance. Your evaluation of the RESCUER App should be based on your personal experience.
- It is important to remember that this is not an evaluation of you personally. Rather, we are interested in your personal evaluation of the RESCUER App
- When you do not have any comment to an open question, please, cross the lines for the answer so that I know that you did not answer on purpose.

8.



a) What do you like about this screen?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

b) What do you not like about this screen/ What should be improved?

_____

_____

_____

_____

_____

_____

_____

f) Do you have any other comments about this screen?

_____

_____

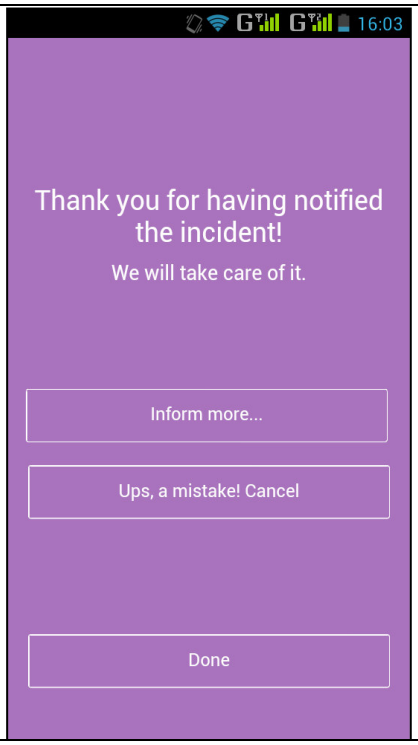
_____

_____

_____

_____

_____



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

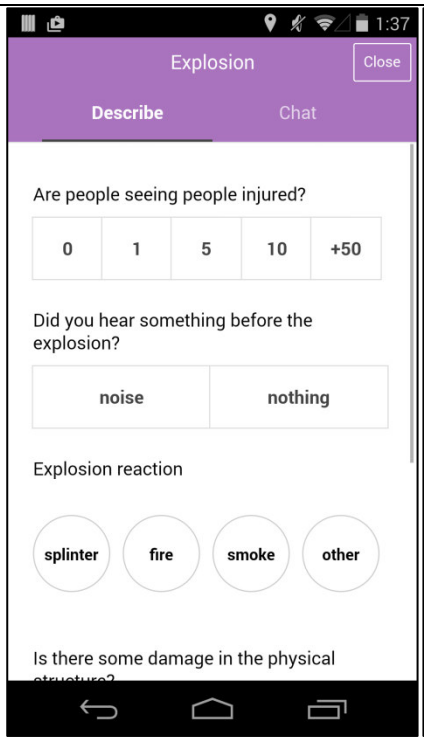
_____

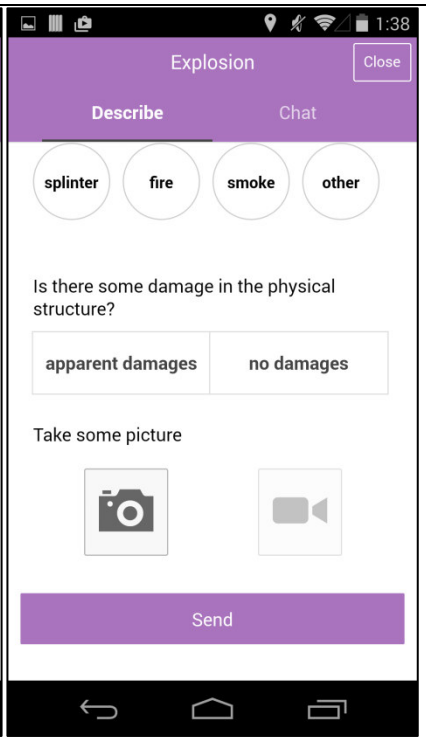
c) Do you have any other comments?

_____

_____

9.





Upper part of the screen lower part (when scrolling down)

t) What do you like about this screen?

_____

_____

_____

_____

_____

_____

u) What do you not like about this screen/ What should be improved?

_____

_____

_____

_____

_____

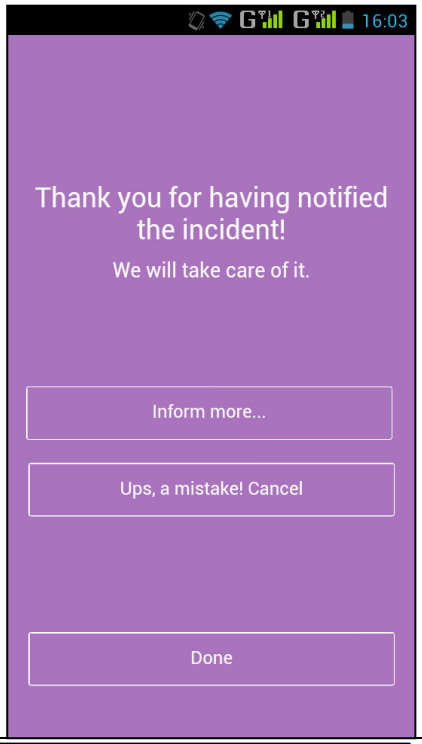
_____

_____

v) The questions at the screen are

clear O O O O O unclear

	w) Do you have any other comments about the questions in the app? _____ _____ _____ _____ x) It is clear that I have to click on the send button to send my data: O yes O no y) The answer possibilities are in the app clear O O O O O unclear z) Do you have any other comments about the answer possibilities? _____ _____ _____ _____ _____ aa) Do you have any other comments about this screen? _____ _____ _____ _____ _____



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

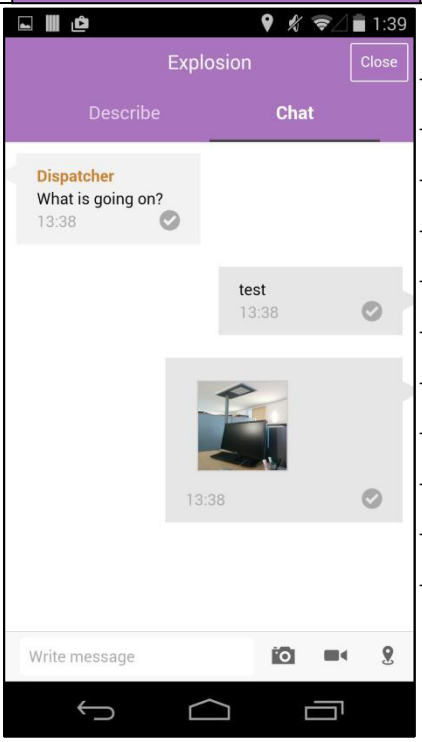
_____

c) Do you have any other comments?

_____

_____

10.



a) What do you like about this screen of the chat function?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

a) What do you not like about this screen?/ What should be improved?

_____

_____

_____

_____

_____

_____

_____

b) Do you have any other comments about this screen of the chat function?

_____

_____

_____

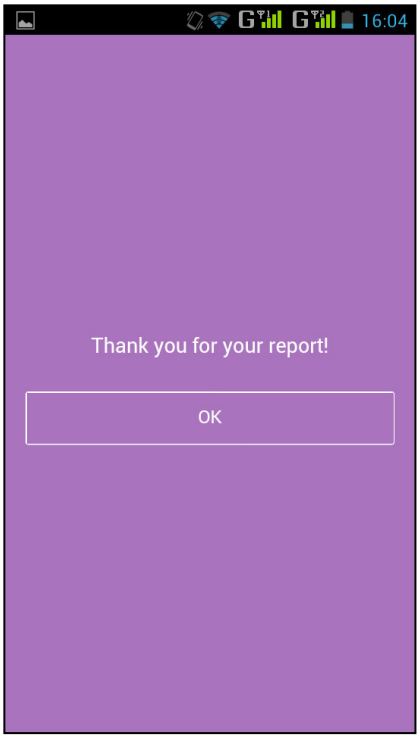
_____

_____

_____

_____

11.



a) What do you like about this screen?

_____

_____

_____

_____

b) What do you not like about this screen? What should be improved?

_____

_____

_____

_____

c) Do you have any other comments?

_____

_____

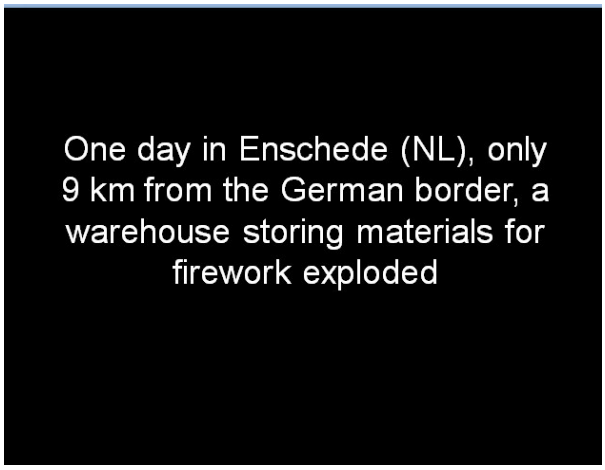
The RESCUER App…	--- -- - -/+ + ++ +++	**The RESCUER App…**
requires unnecessary input.	O O O O O O O	does not require unnecessary input.
provides insufficient information about which entries are valid and necessary.	O O O O O O O	provides sufficient information about which entries are valid and necessary.
complicates orientation due to an inconsistent design	O O O O O O O	facilitates orientation due to a consistent design.

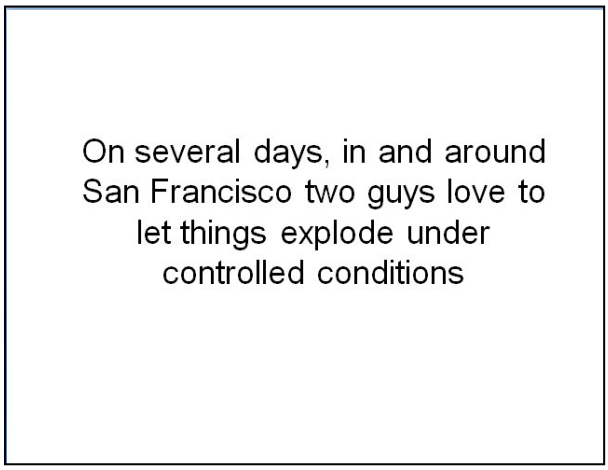
The RESCUER App…	--- -- - -/+ + ++ +++	**The RESCUER App…**
requires the memorization of too many details.	O O O O O O O	does not require the memorization of too many details.
requires a lot of time to learn	O O O O O O O	requires little time to learn.
is difficult to learn without external support or a handbook	O O O O O O O	is easy to learn without external support or a handbook.

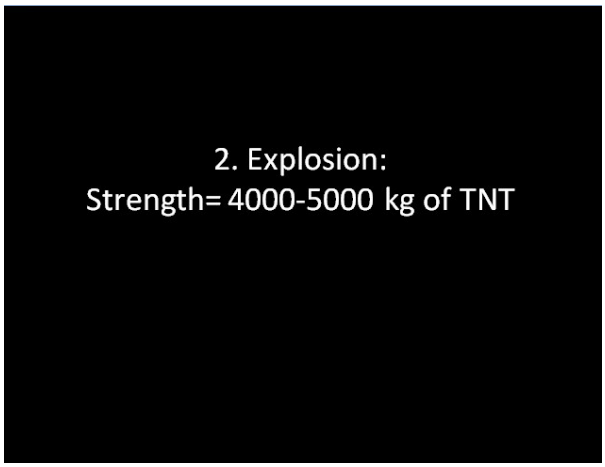
12.	Would you use the app in a real emergency? O yes, because _____ O no, because_____ O I don't know, because _____
13	What do you think is the purpose of the experiment? _____
14.	Do you have any other comments about the experiment or the app? _____ _____ _____ _____ _____

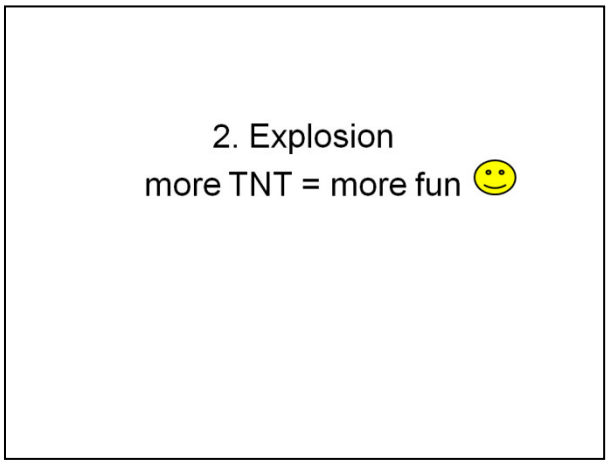
Appendix H. Example Slides of the Presentation about an Explosion

The slides on the left belong to the experimental group. The one on the right present the slides of the control group. Before the presentation, it was shortly explained who the Mythbusters are.The slides are meant to give an impression of the presentation. The pictures were derived from YouTube videos and pictures on the internet.









The explosion was so strong that
200 flats were totally destroyed,
resulting in 1250 people left homeless,
windows were destroyed in a radius of 1,5 km

The explosion was so strong that the whole car was destroyed, resulting in a happy Mythbusters-team.
Pieces of the car were spread over a radius of 50 m

23 persons died

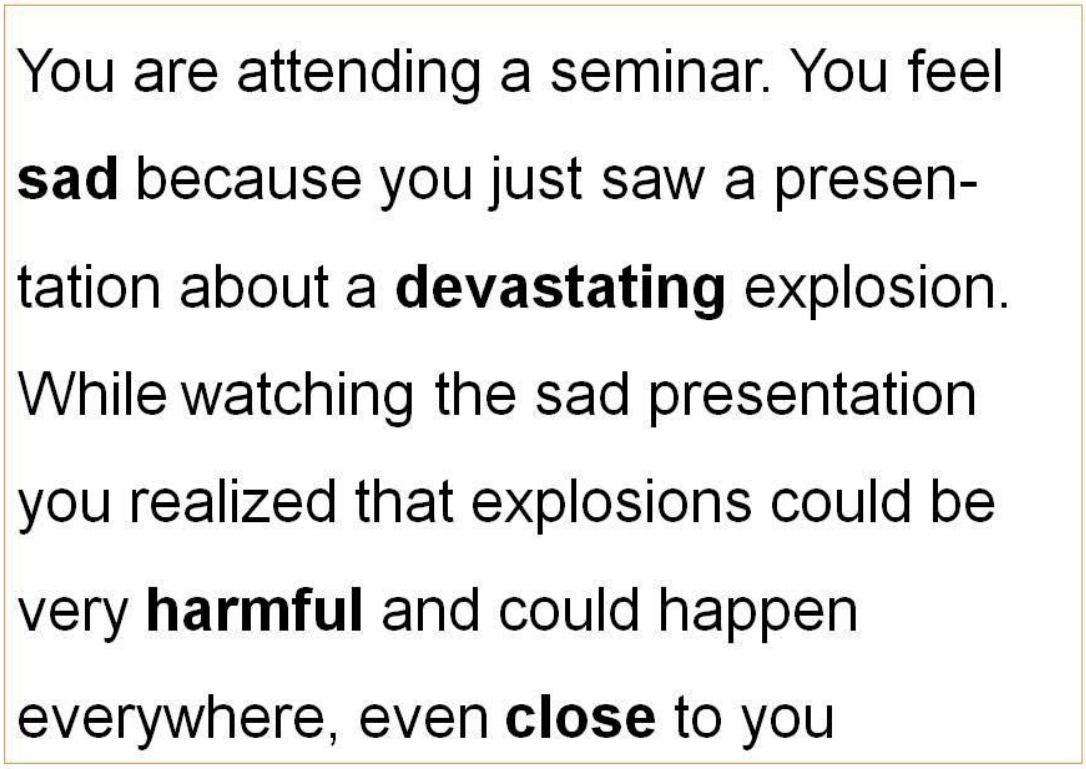
But the car died


TOYS
TOYS
TOYS
TOYS

Appendix I. Text in Scenario

The text was presented in form of a PowerPoint presentation with fixed presentation times. This ensured that the presentation time was the same than in the experimental and control group. The first three slides present the text of the experimental group. Thereafter, the slides belonging to the control group are presented.

Experimental Group:



This makes you feel **afraid**. You have a sense of foreboding that something **bad** will happen soon. You cannot explain why you feel this way. Therefore, you feel **insecure** and nervous. You can see in the faces of the other people in the room that they feel the same.

Seeing them being **nervous** as well makes you feel even **worse**. You look out of the window, to check whether you see anything that would explain your sense of foreboding. But you only see these two guys on a street:

Control group:

You are attending a seminar right now. You are in a **good** mood after watching a **funny** presentation about the Mythbusters. While watching the presentation, you realized that explosions can be quite **fun**. You know that Mythbusters cooperate with the research institute DFKI, which is about 200 meters **distance** from Fraunhofer.

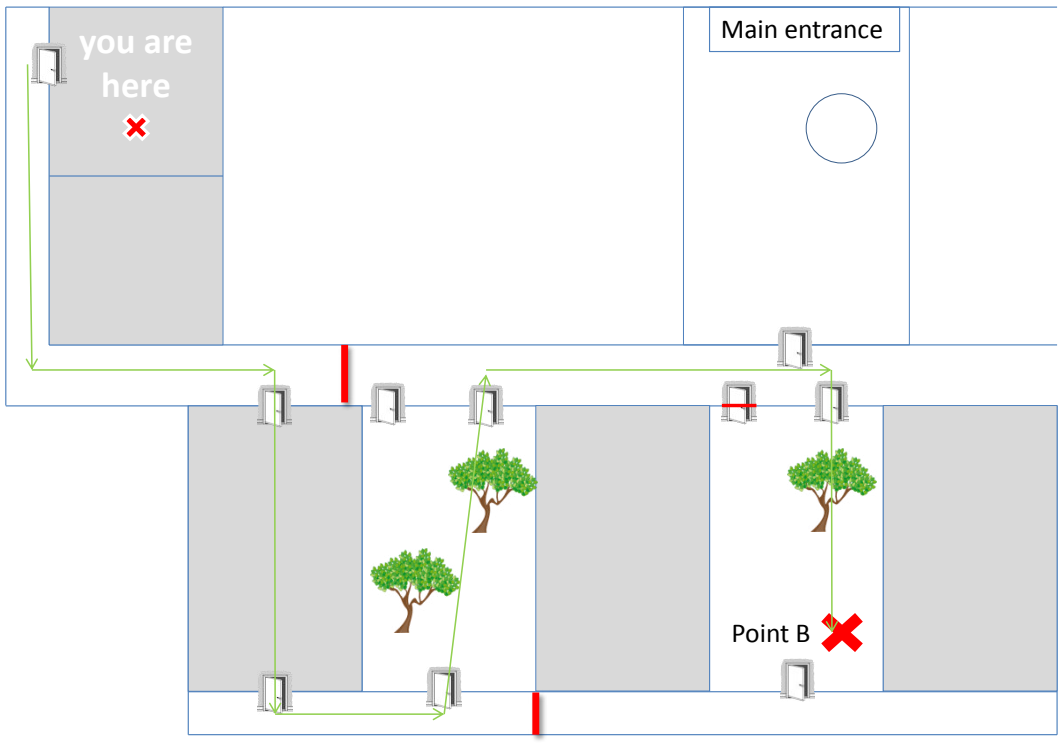
They will conduct an experiment today. You have heard that some of their **scientific** experiments are not going the way they planned them. Mostly the experiments that go wrong are **funny**. You know that you are in a **safe** distance to DFKI, so you are safe in case that the experiment with DFKI goes a bit wrong.

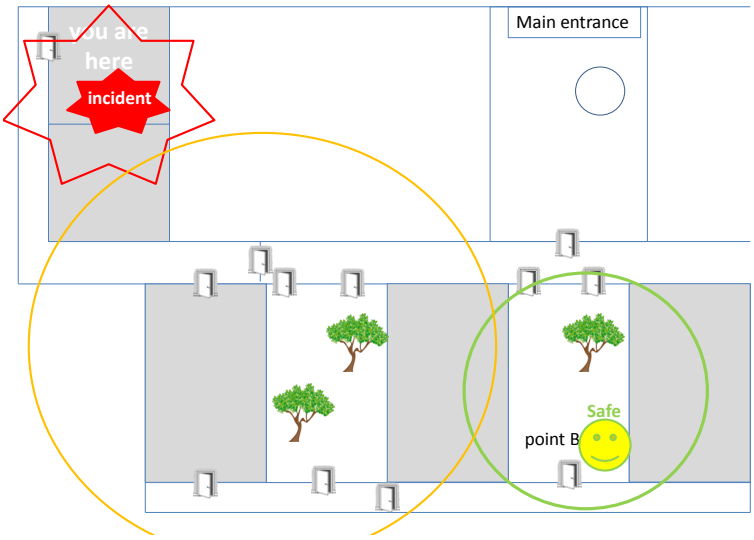
Right now there is a break at the seminar. You look out of the window and you see that the experiment in **cooperation** with DFKI indeed ended slighly unexpected:

Appendix J. Instructions for Assistants including Maps of the Location

This document was handed out to the assistants so that they were informed what exactly will happen during the experiment.

This document provides you with information and instructions.
The experiment will take place two times on Tuesday (17th of November).
One time from 10:00-11:15 am and one time from 3-4:15pm.
The following paragraph will introduce you to the experiment. Please, do NOT share any of this information with anyone. It is very important that participants do not know about this information. Also do not tell the information to people who are not participating (you know how it is going, the person tells someone else, and this person tells someone else who is maybe a participant)
The experiment is about testing the RESCUER App in a simulated emergency situation. An incident is simulated in Raum Koblenz. Point B (Atrium btw. A and B, see the picture below) is supposed to be a meeting point. Point B represents safety.





The experiment requires that people walk from Raum Koblenz to point B. The participants of the experiment have to walk the route depicted below. Your job is to point in a direction and tell the people that they should walk in this direction. There are in total 6 persons who indicate the direction to the participants.

Control group:

The group in the morning (10-11:15 am) is the control group. The participants should be relaxed and NOT experience any stress. Your task is to explain them the way with a friendly voice and a smile =) The control group is informed about the route. You are there to make them feel comfortable by reminding them about the direction in a friendly and calm way. Smile =)

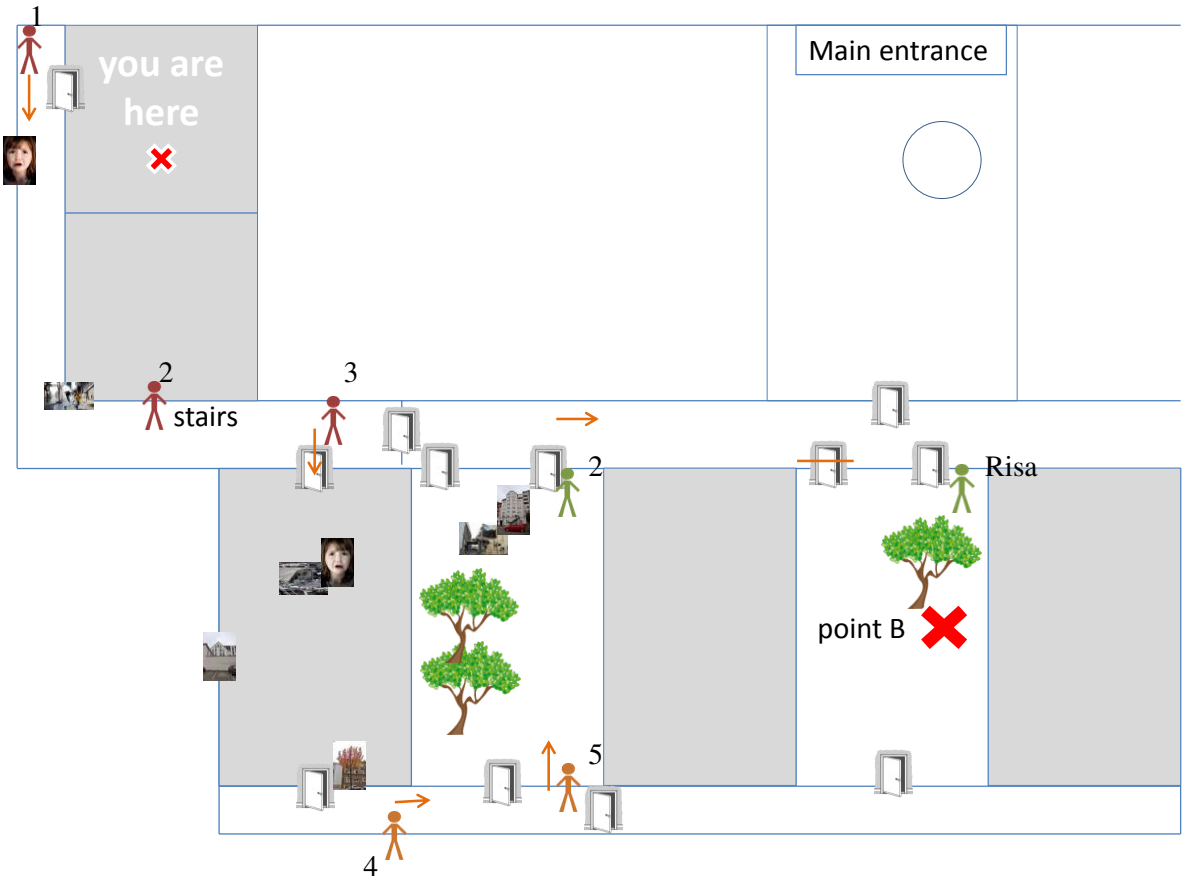
Experimental group:

The group in the afternoon (3-4:15 pm) is the experimental group. The participants are supposed to experience stress. Your task is to tell them the way while making them feel stressed. The experimental group is NOT informed about the route in advance. You have to tell them the direction in a stressful way. Imagine, an emergency has happened and you have to tell the people were they have to go and that they should hurry. Imagine, you are stressed yourself by the situation and are behaving stressed and unfriendly to those people. Behave like this. Speak loudly, look stressed, act hectic.

In total, you have to complete four tasks. 1. Tell the direction 2. Act relaxed/ stressed 3. Keep the doors open 4. In the experimental group: take care of the people (see 'instructions')

Instructions:

- Do NOT mention the word "run". Use words, such as "hurry" and "go"
- In the experimental group: when you see people who give you the impression of being extremely stressed. Do not stress them any further -> act calmly. If you have the impression that the person is about to panic/ behaving totally confused/ without control over themselves: Tell, the person to stop with the experiment!
- Keep doors open! Closed doors increase the risk of people getting hurt
- Tell persons who are not participating but accidently passing by to not cross the way of the participants. You can explain to them where the participants are supposed to walk along.

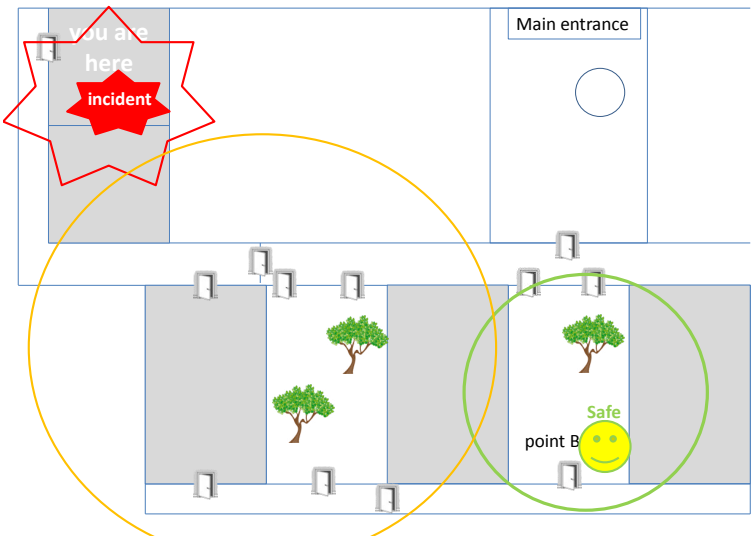


[The final experiment differs a bit from this representation. There were no photos located between person 1 and 2. Peroson 2 stood a bit more to the left to keep the door of the corridor open. The participants entered through a third door and went directly into the meeting room instead of going to the position marked with a red cross]

The location of you and other people supporting me is the same in both groups. In the control group all of you are acting the same. But in the experimental group, you have to induce different levels of stress. Red means much stress, orange medium stress, and green no stress.

	Location	
Person 1:	next to door of room Koblenz	hurry, go to point B, this way
Person 2:	in front of staircase	hurry, go to point B, this way
Person 3:	entrence of C.0.	hurry, go to point B, this way
Alberto:	behind C.0. at staircase	go this way
Guillermo:	in atrium btw. Block C and B	go this way
Person 2:	keeps door open in this atrium	go to the right
Person at Point B:	in front of room Mainz (atrium btw. A and B)	fill in last information in app, if you like to. Then, take a seat in this room.

Person 4 also has to turn the music on in advance of the experiment.
Person 3 has one more task, namely closing the door to C.0 and turning off the sound of fire alarm after the last person past person 3.

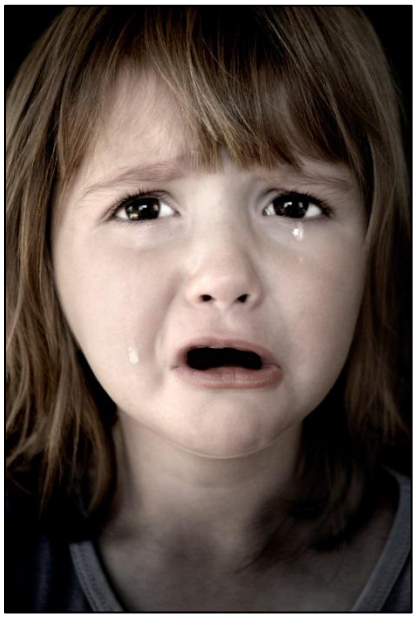


[Description of the zones (red=hot zone= immediate danger, orange= warm zone= danger, green= cold zone=safe)]

Appendix K. Photos on Posters

This document shows an exception of photos that were printed as posters in approximately A2 format. The photos depicting negative stimuli were visual emotional stressors. The posters remaining of the participants´ task were printed in A3 format. The posters hang along the path to point B. The black picture boarders were not part of the posters but serve to better illustrate the photos in this document.

Negative emotional stimuli for the experimental group:



Photos of destroyed Buildings:

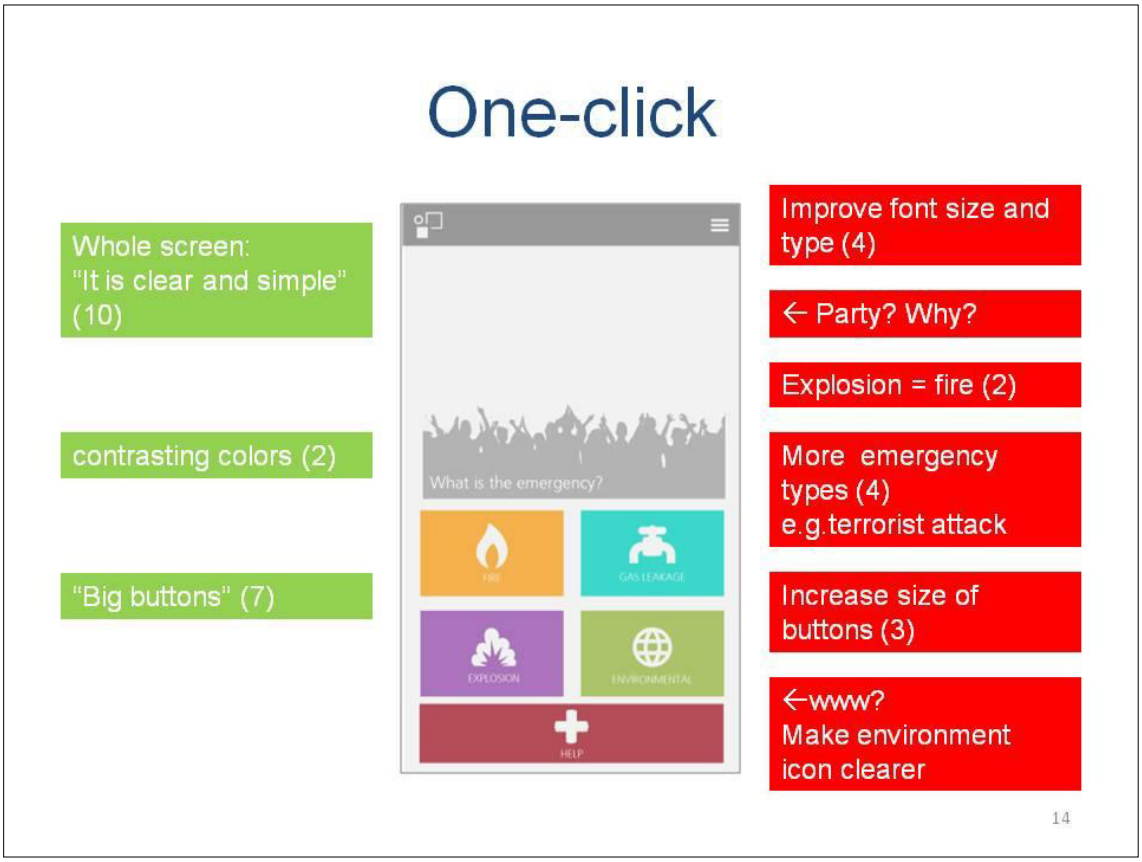
Neutral stimuli in both groups:

Appendix L. Powerpoint Presentation of Results with regard to the RESCUER App

The presentation was given to the two project leaders and members of the RESCUER project. The slides about the stress manipulation are removed since those results are already discussed in the chapter 'results' in this thesis.

The number in brackets represent how often a comment was made. The boxes with green background present what the participants liked about the screen, the red boxes present what they do not like about the screen and/ or what should be improved. During the presentation additional information was presented, for instance, to which group in the experiment the writer of a critique belongs.

Some information was only given verbally. On slide 15, it was explained that 5 persons indicated that it was not clear that they have to use a send button to send information. On the slide, only the design recommendation is presented, namely that there should be no scrolling, so that the button is not missed.



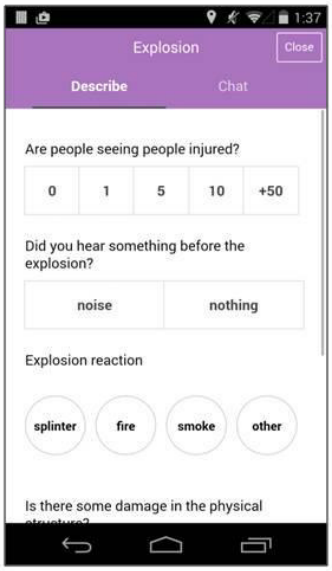
Guided interaction

"shows good options"

"simple choices, not hard to make a decision"

"Taking pictures or video is really good option" (2)

"The tabs describe and chat" (1)



Takes too much time

Formulation of questions is not clear (question 1 & 4)

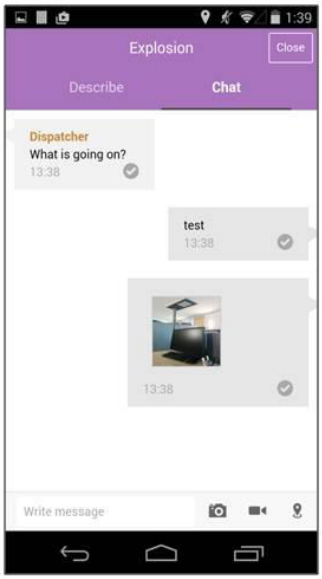
Request:
- No scrolling
-upload photos from device
-Photos first
-More than 1 picture

15

Chat

Familarity: "resembles the whatsapp chat screen " (5)

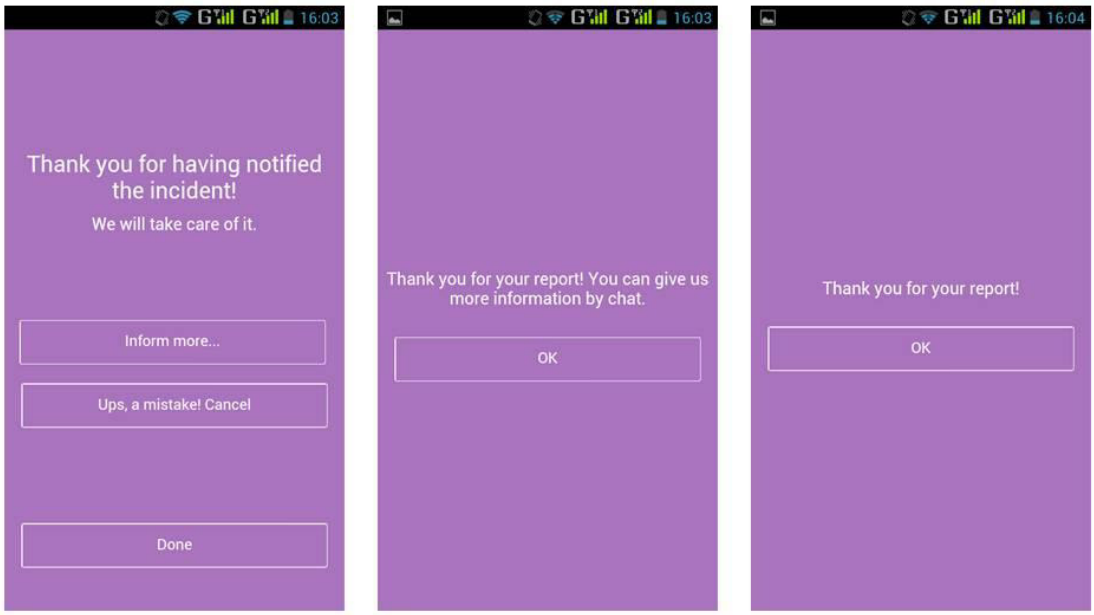
"simple and clear "



„als ich später zufällig auf close gedrückt habe, konnte ich den chat nicht mehr öffnen =("

16

Confirmation screens



17

Confirmation screen

ExpMit:"Its simple and easy to understand during threatful situations "(1)	"there are a lot of text. I did not read during the experiment." (in expMit 4 /6)
"gives opportunity to take report back"	Spelling of oops
"It is good to get an	ScreenT3: this screen not required
"It reminds of presence of the "chat " tab in the app"	ExpOhne: The report should not be ended here
	Buttons are difficult to read (6)
ExpMit: "A happy image could be better, or some image which could bring down the stress level."	"reported like " Thank you for reporting the explosion incident, We will take it in priority" "

18

What is the emergency?		
explosion	fire	Explosion & fire
*****		
***		
****	*	
**		****

Are people seeing people injured?			
0	1	5	Not answered
***	*		*
	*	**	
	***	**	
	****		**

Did you hear something before the explosion?		
noise	**nothing**	Not answered
*	****	
**	*	
**	***	
***	**	*

Explosion reaction						
splinter	smoke	fire	Splinter + smoke	**Splinter + fire +smoke**	Fire + smoke	Not answered
		*	*	*		**
**						*
*	**	*	*			
	**				**	**

Is there some damage in the physical structure?		
Apparent damage	No damages	Not answered
*****		
**		*
***	**	
****		**

Control Without
Control With
Experim. Without
Experim. With

19

Chat function

Control:

"saw an explosion, think it was a car bomb"

"I also saw people around danger zone that seems they are not aware of danger".

Experimental:

"Two people are walking on the street",

"Ther were cars and houses in the area",

"(…) the one car exploded"

"(…) resulted in killing of people, many of them got injured"

20

Chat function

Control:
"**saw** an explosion, **think** it was a car bomb"
"I also **saw** people around danger zone that **seems** they are not aware of danger".
-> distinction btw. saw and think-> more objective

Experimental:
"Two people are walking on the street",
"Ther were cars and houses in the area"
-> before explosion: quite objective

"(...) the one car exploded"
"(...) resulted in killing of people, many of them got injured"
-> description of explosion and situation after it:
Reporting what they think has happend (wrong information)
-> not objective

21

Acceptance/ Understanding

Answers that express that the purpose of the app is not clear:

- But, nobody will chat during an emergency. I will chat only if am 1-2kms away from blast, after at least 15 mins into safe zone
- No one will have time to chat while trying to get to a safe place. Only useful when one is already safe.
- Personally I prefer to talk in cas of emergency. I dont think in a real emergency situation people tend to send message and wait for answer
- calling can be added!
- Better if its 24x7 online chat facility
- Some more options can be added to this screen, like car/train accident
- An emergency number could be better
- However, in the chat window, there should always be one person available
- Is not nessecary no body want to chat in urgent

22

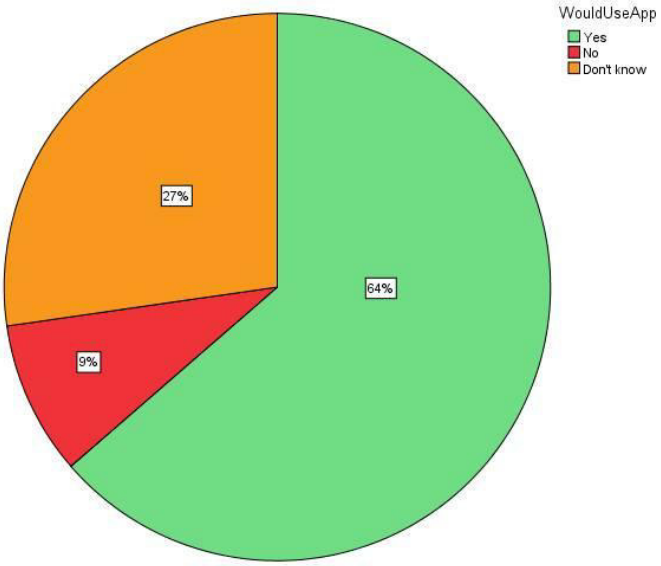
Acceptance/ Understanding

Using app is intuitive ≠ purpose of app is intuitive

-> could harm the **acceptance** of the app & could cause the sending of **irrelevant** information through the app

23

Would you use the app?



24

Would you use the app?

- using phone during emergency ?
- Purpose of app not clear
- No direct experience with app

- location
- Emergency service arrives faster
- Warning people
- Faster than calling
- Pictures

- Calling is easier
- internet connection

25

Summary

Positive:

- Intuitive use of app
- No comments about severe problems
- Most people would use app

Improve:

- People do not answer objectively
  -> carefully analyze reported information
- Purpose of app is not clear
  -> detailed demonstration of app, let people explore app

26

The text below summarises the most import findings regarding the app.

Summary of the most important findings regarding the app

The RESCUER App has been evaluated several times. One evaluation conducted in a laboratory setting had tested how well people were able to use the app under stress (described in detail in chapter 4.4 of Deliverable D2.1.3). Stress was induced with a one-back task, causing cognitive load. This evaluation was already a good start to test the app under fairly realistic circumstances, but a laboratory study can never truly reflect reality. Therefore, another experiment was conducted with the objective to evaluate the app in a more realistic setting, but within the ethical constraints of social research. For this purpose, an emergency situation was simulated. The emergency type was an explosion. In a real use context, the app users are moving to a safe place. Therefore, the participants had to get from one room to another room via hallways and an atrium inside the building. The distance could be covered by walking within three minutes. The participants also had to report information they saw in a video prior to fleeing, and on pictures that were along the path to the second room. Some participants were not provided with a smartphone to use the app, since not every person in a real emergency situation has a smartphone with this app installed. Those participants without a smartphone were asked to report their observations afterwards by filling out a questionnaire that covered all questions that the app also asked. The information that was reported through the app and the questionnaire were analysed, as well as the information that was obtained by a further questionnaire. In this questionnaire, the participants of both groups could express their opinion about the different screens of the app.

The participants reported that the app is intuitively to use. No severe problems with using the app were reported. The one-click interaction screen was rated quite positively, except that some participants

did not report fire as an explosion reaction, but as an emergency type. The guided interaction screen was criticized in terms of language and understandability, especially the questions "Are people seeing people injured?" and "Is there some damage in the physical structure?". The answers to all questions asked by the app or questionnaire varied widely. For instance, the question about the number of injured persons varies from zero to several. However, none of the experiment's materials depicted even a single injured person. Similar findings were found in previous studies (see Deliverable D2.1.3, chapter 4.4). Most participants appreciate the chat function, since it looks similar to other chat apps. Some chat messages indicate that the participants who experienced stress reported what they *think* has happened instead of describing the situation objectively. To sum it up, the screens are fine, and only small changes should be made. However, the reported information has to be interpreted with care.

Several comments to the screens, such as "calling could be added", indicate that some participants had a wrong understanding of the app. Using the app may be intuitive, but that does not mean that the purpose of the app is self-evident. The app and its purpose were presented before the experiment started. The introduction that guided the participants through the report of a fire may not have been detailed enough. One misunderstanding already became obvious in a pre-test. The one-click interaction screen was perceived as a menu button but not as a button to directly send first information about the emergency. Many misunderstandings revolved around the chat function. It was not clear to some participants that the chat function is not intended to be used in the hot or warm zone (even if this was discussed in the introduction).

Still, most participants would use the app in a real emergency (64%), some were not sure (27%), and only 9% would not use the app. The reasons for using the app were the use of pictures in reporting, easily reporting the location, warning other people, and that using the app is faster than calling, resulting in a faster arrival of the emergency services. Reasons for not using the app were the opinion that calling is faster, and the risk of a poor internet connection. A reason for being unsure was among others that the participants did not have direct experience with the app.

To sum it up, the RESCUER App is already very good. It could be used intuitively, however, the purpose of the app is not self-evident. This could harm the acceptance of the app.

Recommendations:

- Before first use, the users should be provided with a few introduction screens that enable them to:
  - receive a detailed demonstration of the app
  - receive an explanation of the purpose of the app and the intentions of the screen
  - have the opportunity to explore the app
- One-click interaction screen:
  - Remove the picture above the buttons. The picture confuses users.
- Guided interaction screen:
  - The questions have to be formulated clearer.
  - It should be possible to send more than one photo.
  - Scrolling should not be required, as this may cause users to overlook the send button.
- Chat screen:
  - There should be the possibility to reopen the chat function when it was closed accidentally.
- Confirmation screens:
  - The text should be shortened. Users under stress do not have the time to read it.

- The RESCUER system should instruct and remind the emergency services to interpret the reported data critically.

Appendix M. Result Tables

Descriptive Statistics

Table A

Descriptive Statistics of the Stressfulness of the Experiment Elements Rated by the Whole Experimental Group Ordered from Most to Least Stressful

	N	Mean	Std. Deviation
Sound of Smoke Detector	12	3.42	1.08
Video of Explosion	12	3.42	0.90
Using the App	7	3.29	0.76
Assistants	12	3.17	0.94
Walk to Point B	12	3.08	1.00
Presentation about Explosion	12	3.00	1.13
Finding Way	12	2.75	1.06
Pictures of Humans	11	2.73	0.90
Pictures of Explosion	11	2.64	0.92
Sound of Screaming People	11	2.55	0.93
Persons in the crowd	12	2.08	0.90
Valid N (listwise)	6		

Table B

Descriptive Statistics of the SAM Scales Sorted by Control and Experimental Group

		Difference between rating during and before the experiment		Rating during the experiment	
		M	SD	M	SD
Control	Valence	-1.00	2.21	5.20	1.55
	Arousal	1.80	1.93	6.00	1.05
	Dominance	-1.10	2.32	5.40	1.90
Experimental	Valence	-2.08	1.31	4.00	0.95
	Arousal	1.50	2.28	6.33	1.56
	Dominance	-1.33	1.44	4.25	1.48

Table D

Mean and Standard Deviation of the SAM Scales Sorted by Group and App

			Difference between before and during			During the experiment	
Group	App	SAM Scale	N	M	SD	M	SD
Control	Without	Arousal	5	1.80	1.30	5.80	0.45
		Valence	5	-0.20	1.30	5.20	1.30
		Dominance	5	-0.40	1.82	4.80	2.17
		Valid N (listwise)	5				
	With	Arousal	5	1.80	2.59	6.20	1.48
		Valence	5	-1.80	2.77	5.20	1.92
		Dominance	5	-1.80	2.59	6.00	1.58
		Valid N (listwise)	5				
Experimental	Without	Arousal	5	1.00	3.32	5.80	1.64
		Valence	5	-2.40	1.34	3.80	0 .84
		Dominance	5	-1.80	1.10	4.60	0.89
		Valid N (listwise)	5				
	With	Arousal	7	1.86	1.35	6.71	1.50
		Valence	7	-1.86	1.35	4.14	1.07
		Dominance	7	-1.00	1.63	4.00	1.83
		Valid N (listwise)	7				

Table E

Means and Standard Deviation of the Subjective Stress Ratings Sorted by Control and Experimental Group and Separately Sorted by App

	Difference between during and before the experiment		Subjective stress rating during the experiment	
	M	SD	M	SD
Control	3.70	0.63	4.90	0.59
Experimental	4.17	0.60	5.67	0.62
With App	4.17	0.37	5.92	0.26
Without App	3.70	0.86	4.60	0.86

Table F

Mean and Standard Deviations of the Subjective Stress Ratings Sorted by Control and Experimental Group and Users Respectively Non-Users of App

Groups		Difference between measure 1 and 2		Subjective stress rating at measure 2	
Group	App	Mean	Standard deviation	Mean	Standard deviation
Control	Without	2.80	2.17	3.60	1.67
	With	4.60	1.52	6.20	0.84
Experimental	Without	4.60	3.13	5.60	3.36
	With	3.86	1.07	5.71	0.95

Note. Measure 1= before the experiment, measure 2= during the experiment

Tables of Fixed Effects

Table G

Type III Tests of Fixed Effects of the Dominance Ratings

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	19.64	523.27	.00
moment_bin	1	40.00	15.63	.00
smartphone	1	19.64	1.45	.24
condition	1	19.64	4.68	.04
moment_bin * smartphone	1	40.00	.84	.37
moment_bin * condition	1	40.00	.06	.81
smartphone * condition	1	19.64	5.64	.03
moment_bin * smartphone * condition	1	40.00	.60	.44

Table H

Type III Tests of Fixed Effects of the Arousal Ratings

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	20.30	574.06	.00
moment_bin	1	40.00	15.20	.00
smartphone	1	20.30	.79	.38
condition	1	20.30	.79	.38
moment_bin * smartphone	1	40.00	.56	.46
moment_bin * condition	1	40.00	.20	.66
smartphone * condition	1	20.30	.12	.73
moment_bin * smartphone * condition	1	40.00	.09	.77

Table I

Type III Tests of Fixed Effects of the Valence Ratings

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	21.42	965.10	.00
moment_bin	1	40.00	18.50	.00
smartphone	1	21.42	.72	.40
condition	1	21.42	6.62	.02
moment_bin * smartphone	1	40.00	.15	.70
moment_bin * condition	1	40.00	1.54	.22
smartphone * condition	1	21.42	.41	.53
moment_bin * smartphone * condition	1	40.00	1.72	.20

Table J

Type III Tests of Fixed Effects of the Stress Ratings

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	21.74	186.34	.00
moment_bin	1	40.00	22.87	.00
smartphone	1	21.74	4.94	.04
condition	1	21.74	1.44	.24
moment_bin * smartphone	1	40.00	.01	.92
moment_bin * condition	1	40.00	.01	.92
smartphone * condition	1	21.74	1.85	.19
moment_bin * smartphone * condition	1	40.00	.70	.41

Table K

Fixed Effects of Cortisol Values without Extreme Values

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	19.95	74.02	<.01
Moment	1	19.95	2.14	.16
Group	1	19.95	3.73	.07
App	1	19.95	5.55	.03
Moment * Group	1	19.95	0.02	.89
Moment * App	1	19.95	0.45	.51
Group * App	1	19.95	2.63	.12
Moment * Group * App	1	19.95	<.01	.98

Estimates of Covariance Parameters

The variance of the participants is smaller than the variance of the residuals. Moreover, the standard deviation (calculated out of the variance) is relative small when compared to the main effects. That means, the variance between the participants is not dramatical.

Table L

Estimates of Covariance Parameters of the Stress Ratings

Parameter		Estimate	Std. Error
Residual		4.094643	.915590
Intercept [subject = ParticipantID]	Variance	.315542	.637890

Table M

Estimates of Covariance Parameters of the Valence Ratings

Parameter		Estimate	Std. Error
Residual		1.240952	.277485
Intercept [subject = ParticipantID]	Variance	.142257	.207105

Table N

Estimates of Covariance Parameters of the Arousal Ratings

Parameter		Estimate	Std. Error
Residual		1.596667	.357026
Intercept [subject = ParticipantID]	Variance	.527390	.372714

Table O

Estimates of Covariance Parameters of the Dominance Ratings

Parameter		Estimate	Std. Error
Residual		1.260476	.281851
Intercept [subject = ParticipantID]	Variance	.743510	.399105

Appendix N. Testing of Assumptions

This document presents the testing of the assumptions of the statistical models.

The normality of the residuals was tested by conducting an exploratory analysis, including testing of normality. The syntax for all variables looks alike. Therefore, the syntax is only presented for a single case.

Example syntax of testing the assumptions:

```
GET
  FILE='D:\DocumentsD\DataExp\DataStress.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
EXAMINE VARIABLES=Residuals_Stress_LMM BY condition
  /PLOT BOXPLOT STEMLEAF
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Example syntax of saving the residuals of the SAM scales:

```
COMPUTE moment_bin=1 * (moment = 2).
EXECUTE.

MIXED SAM_Arousal BY smartphone condition moment
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=smartphone condition moment smartphone*condition smartphone*moment condition*moment
    smartphone*condition*moment | SSTYPE(3)
  /METHOD=REML
  /RANDOM=INTERCEPT | SUBJECT(ParticipantID) COVTYPE(VC)
  /REPEATED=moment | SUBJECT(ParticipantID) COVTYPE(DIAG)
  /SAVE=RESID.
```

Normality is tested by the Kolmogorov-Smirnov test. The residuals are normality distributed if the test is not significant. A significance level of .05 is applied.

Cortisol

The residuals were obtained from a LMM analysis. LMM is applied because it is applicable for repeated measures, fits better to small samples than an ANOVA, and with more than one factor variable can be put into the model.

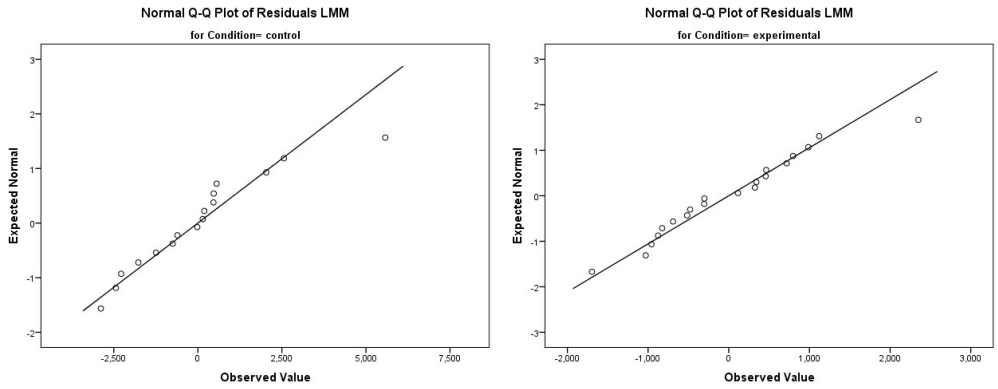
Table b

Tests of Normality of the Cortisol values

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	control	.209	16	.060	.918	16	.157
	experimental	.126	20	.200*	.967	20	.689

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



➔ The cortisol measures are normally distributed. The LMM will be applied.

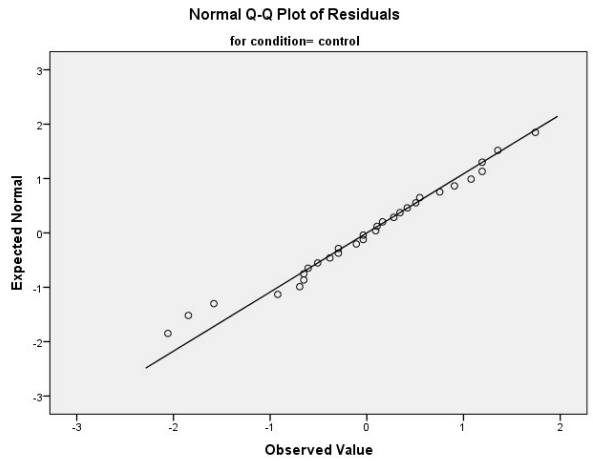
Arousal:

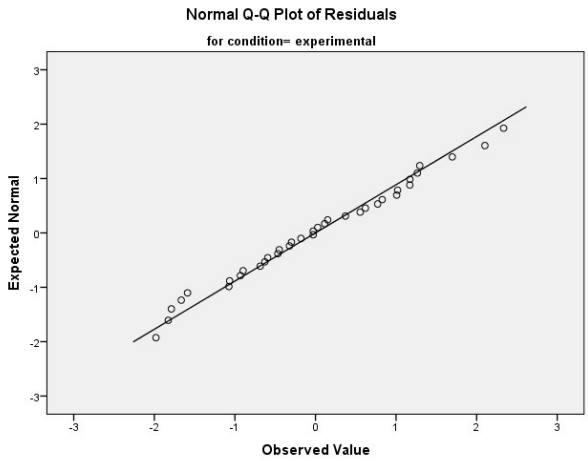
Tests of Normality of the Arousal Ratings

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	condition	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	control	.092	30	.200*	.978	30	.762
	experimental	.064	36	.200*	.979	36	.695

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction





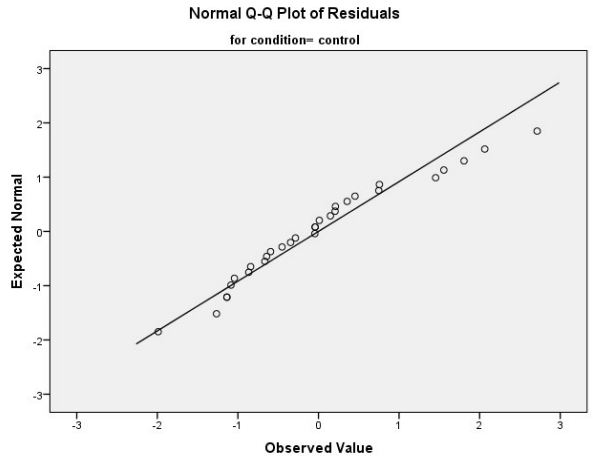
Valence:

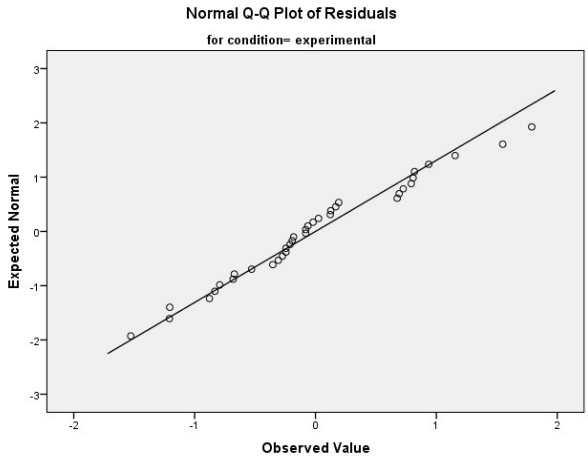
Tests of Normality of the Valence Ratings

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	condition	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	control	.123	30	.200*	.953	30	.207
	experimental	.123	36	.186	.975	36	.587

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction





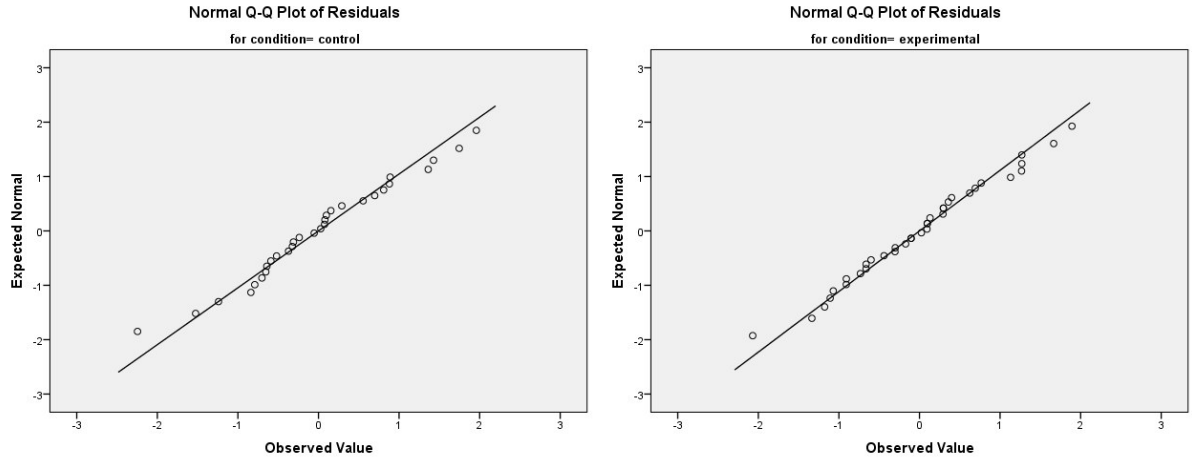
Dominance:

Tests of Normality of the Dominance Ratings

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	condition	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	control	.103	30	.200*	.982	30	.879
	experimental	.078	36	.200*	.987	36	.941

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



Subjective Stress ratings:

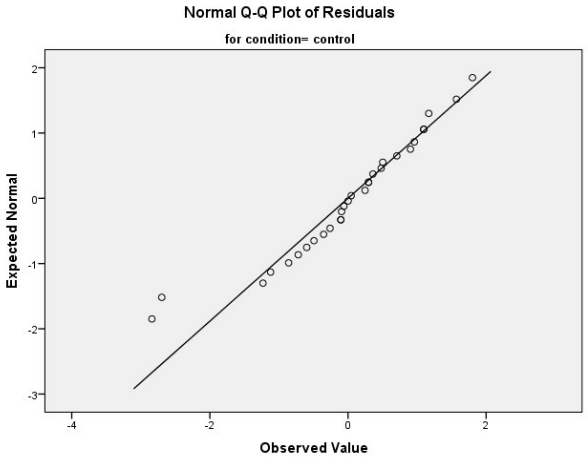
Tests of Normality of the Subjective Stress Ratings

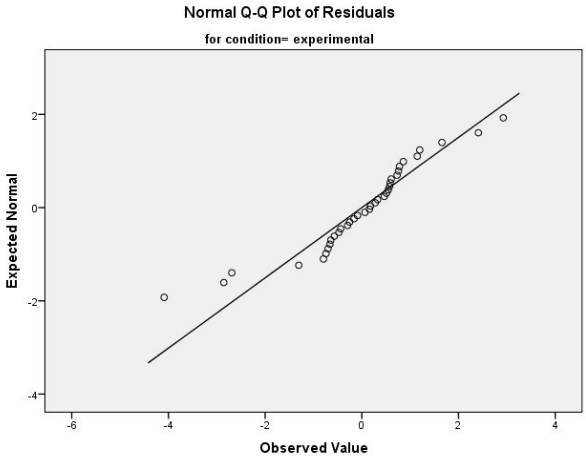
		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	condition	Statistic	df	Sig.	Statistic	df	Sig.
Residuals	control	.128	30	.200*	.935	30	.066
	experimental	.164	36	.016	.921	36	.014

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

➔ In the experimental group were the residuals NOT normally distributed


Normal Q-Q Plot of Residuals
for condition= control
Expected Normal
Observed Value
2
1
0
-1
-2
-3
-4
-2
0
2


Normal Q-Q Plot of Residuals
for condition= experimental
Expected Normal
Observed Value
2
0
-2
-4
-6
-4
-2
0
2
4

Appendix O. Syntax

This document presents the syntax of the statistical analysis. The syntax of the descriptive statistic tables are not presented since the configuration of those is obvious. Some names of the variables were changed in the tables; Smartphone= app, Condition= Group.

Syntax of the results

Cortisol LMM exluding 4 participants because of extreme values

```
MIXED Cortisol BY Moment Condition Smartphone
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=Moment Condition Smartphone Moment*Condition Moment*Smartphone
Condition*Smartphone
    Moment*Condition*Smartphone | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION
  /REPEATED=Moment | SUBJECT(ParticipantID) COVTYPE(DIAG)
  /SAVE=PRED RESID.
```

Example of the syntax of figure 4-8

```
* Chart Builder.
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=moment
MEAN(SAM_Valence)[name="MEAN_SAM_Valence"]
    ParticipantID condition smartphone MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: moment=col(source(s), name("moment"), unit.category())
  DATA: MEAN_SAM_Valence=col(source(s), name("MEAN_SAM_Valence"))
  DATA: ParticipantID=col(source(s), name("ParticipantID"), unit.category())
  DATA: condition=col(source(s), name("condition"), unit.category())
  DATA: smartphone=col(source(s), name("smartphone"), unit.category())
  GUIDE: axis(dim(1), label("moment"))
  GUIDE: axis(dim(2), label("Mean SAM_Valence"))
  GUIDE: axis(dim(3), label("condition"), opposite())
  GUIDE: axis(dim(4), label("smartphone"), opposite())
  GUIDE: legend(aesthetic(aesthetic.color.interior), label("ParticipantID"))
  SCALE: linear(dim(2), include(0))
  SCALE: cat(dim(3), include("1", "2"))
  ELEMENT: line(position(moment*MEAN_SAM_Valence*condition*smartphone),
    color.interior(ParticipantID), missing.wings())
END GPL.
```

Example of the syntax of mixed models as they were applied for the subjective ratings

```
COMPUTE moment_bin=1 * (moment = 2).
```

```
EXECUTE.

MIXED stress BY moment_bin smartphone condition
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=moment_bin smartphone condition moment_bin*smartphone moment_bin*condition
smartphone*condition
    moment_bin*smartphone*condition | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION
  /RANDOM=INTERCEPT | SUBJECT(ParticipantID) COVTYPE(VC).
```