Persuasive System Design Analysis of Mobile Warning Apps for Citizens

Christoph Kotthaus, Thomas Ludwig, Volkmar Pipek
Institute for Information Systems, University of Siegen, Siegen, Germany
{christoph.kotthaus, thomas.ludwig, volkmar.pipek}@uni-siegen.de

Abstract. Large-scale emergencies, such as hurricane Katrina in 2005 that also damaged large parts of New Orleans or the 2013 Central European flood, have shown the importance of an appropriate warning as well as instruction of the affected people on-site. Nowadays modern mobile devices are widely spread throughout the population in many parts of the world and apps are available for warning as well as giving advice, which have the advantage to reach citizens individually, i.e. based on their current location. However, disaster communication is prone to many kinds of biases and strong emotions such as fears, making it difficult to point the crowd in the intended direction. It is therefore all the more important that the messages are well chosen and presented to the users. Comments or feedback from the users about products or services i.e. in web stores, rating web services or app stores are a valuable source for an in-use requirement engineering. Within this paper, we will firstly analyze user comments about two of the most important mobile warning apps in Germany. Secondly, we will correlate the findings with the Persuasive System Design model as a method to design for the domain of emergency management. Based on our analysis we will discuss this approach, revealing that most importantly system trustworthiness and reliability suffer from malfunction of apps and inappropriate messages undermine a successful persuasion strategy.

Keywords: Persuasive technology; persuasive system design model; emergency management; HCI

1 Introduction

When dealing with emergencies, a variety of official organizations, which consist of public authorities with security responsibilities, such as emergency services (e.g. police, firefighters) or public administration is usually involved. As another important actor, but with less engagement in prevention or response strategies, citizens are also engaged in various ways during emergencies [1, 2]. Victims, indirect affected citizens, like family members or neighbors, or volunteers take recovery actions.

One typical characteristic of (large-scale) emergencies, especially regarding mitigation, is that decisions have to be made for low-probability, high-consequence events [3]. This causes well-known biases in human decision making like “[…] the tendency
to learn by excessively focusing on short-term feedback, [...] poor insights into future consequences, [...] and poor inter-temporal tradeoffs between short-term costs and long-term benefits” [3]. Emergency warnings, for example, often prove to be false alarms, as impact zones mostly are much smaller than warning zones, reducing beliefs in related warning messages. Misjudgments regarding future consequences, amongst others, are caused by the subjective assessment of the likelihood that i.e. a hazard will occur [4] and the subjective consideration whether taking mitigation actions will probably prevent future losses [3].

Within these considerations, biases like the availability bias (mental availability of i.e. losses due to a flood or fire), representativeness bias (taking recent history as an implication for long-term likelihoods) [5], optimistic bias (belief that dangerous events will more likely happen to other people than oneself) and projection bias (inability to imagine i.e. one’s home to be destroyed, leading to refuse to evacuate) [6] are well known cognitive dissonances in the application area of (large-scale) emergencies. Further, discrepancies regarding tradeoffs (short-term costs and long-term benefits) underlie biases as well, like the status quo bias (default or no action at all are preferred instead of actions with uncertain outcomes) [7] or the tendency to procrastinate mitigation investments against low-probability events. This is also caused by hyperbolic discounting, the consideration of current relative benefits versus future events [8]. Meyer [3] presents even more biases and causes of misconduct in emergency situation which cannot entirely be discussed here. Due to these biases, citizens are usually often not familiar with concepts of risk communication or warning [9–11].

Large-scale emergencies, such as the 2013 Central European flood or hurricane Katrina in 2005 that also damaged large parts of New Orleans, have shown the importance of appropriate warning as well as risk communication to the affected citizens on-site to overcome possible biased actions. As early forms of warning mechanisms, official organizations used sirens or loudspeakers announcements [12] in combination with radio or television to reach as many citizens locally as possible. However, nowadays modern mobile devices are widely spread throughout the population in many parts of the world. Thus, mobile apps are available for warning as well as giving advice, which have the advantage to reach citizens individually, i.e. based on their current location. As Ludwig et al. [13] have shown by using mobile apps, individual targeted warnings are possible and are more likely to be noticed. However, as Vihamm et al. [14] have shown, institutionally framed warnings are often perceived not well by the public and that citizens “either seek information from informal information networks or simply take their own response action”. Those citizen-initiated actions are not always in line with those of the official organizations. Citizens sometimes enter hazard zones that put themselves in danger or they could interfere with the actions of relief forces [13]. This area of tension leads to the discussion about citizen empowerment version activity control during emergencies: Should citizens be allowed to carry out their activities although these are not in line with the emergency services’? How to manage citizens’ activities without patronizing them and without letting citizens put themselves in danger?

It seems there is a significant need to address citizens in a way to overcome biased behavior. Mobile devices, however, could serve as persuasive technologies, that are
any interactive computing system designed to change people’s attitudes or behaviors […] without using coercion or deception” [15]. Thus, technologies of this kind could be suitable to address these deeply rooted problems. To build persuasive technologies, Oinas-Kukkonen and Harjumaa [16] created the Persuasive System Design model (PSD model). This model allows analyzing and designing systems to be persuasive aiming at closing the gap between the targeted and actual behavior or attitude.

Taking a look at German crisis management, public authorities currently use two mobile apps to warn citizens, namely KATWARN [17] and NINA [18]. Both apps provide functionality to receive warnings, such as weather, flooding, fires or bomb disposals, partly based on the users’ current location. These apps, however, focus on information distribution and general behavioral instructions without deliberately addressing the above mentioned problems.

Within this paper, we will contribute with design implications based on the PSD model [16] to pave the way for overcoming biased actions in emergency situations through technology. To approach this field, we first analyze user comments in Apple iTunes and Google Play Store of the two apps to determine topics and categories. We then apply these findings to the PSD model, starting by analyzing the persuasion context. After that we use the design principals as an anchor to assign the topics and categories. Lastly, this will show areas both apps already fulfill persuasive design requirements as well as reveal respective gaps. This finally leads to our objective to suggest exemplary design implications to specifically expand the design of such apps. We then will outline design implications to make these apps more persuasive.

2 Related Work

Persuasive technology and persuasive system design currently mainly focus application areas like health, environmental sustainability or education. Reducing obesity by promoting individual health behaviors [19], addressing smoking or alcohol abuse [20] or improving responsible gaming [21] aim at closing the gap between actual and targeted behavior or attitude due to the possible gap between short-term satisfaction and long-term consequences regarding diseases like diabetes. Motivating for saving energy [22] or fuel efficient driving behavior [23] are approaches to do this in the area of environmental sustainability, mainly by giving users feedback about their current behavior and the resulting consequences towards the targeted behavior. In education, related work was conducted regarding i.e. study habits amongst students. This was targeted leveraging personal resource management, personal values towards learning and expectations of learning [24].

Little work has been done in the wider context of emergency or hazardous situations regarding persuasive technology so far. Chittaro and Zangrando [25] used persuasive virtual experiences to improve awareness for personal fire safety by simulating dangerous situations to trigger attitude change. Further, technology to persuade visitors during major events to avoid overcrowded places [26] was conducted but without systematically analyzing these using the PSD model.
3 Methodology

3.1 Analyzing user comments

Quite some work was done to analyze user comments, especially with regard to create algorithms to automatically mine topics [27] and determine sentiment, opinion or subjectivity of text messages i.e. shown by [28]. These methods aim at processing large amounts of user comments or reviews to get valuable feedback for the evaluated product, service or app. [29] developed a system to improve quality control of hotels based on user comments from hotel websites. [30] and [31] present an approach to automatically derive software requirements of apps based on user feedback in app stores. Similarly, [32] analyzes comments in Google Play Store with regard to battery consumption of apps. Social networks or social media such as Facebook, Twitter, YouTube, Google+ or Instagram are also subject to opinion mining. I.e. [33] uses a sentiment analysis to determine online radicalization on the online video platform YouTube. Asur and Huberman [34] even predict box-office revenues of movies in advance of their release by analyzing tweets with regard to the sheer number and sentiment. With regard to emergency management [35] created a system to analyze user generated information in various ways, e.g. by its sentiment, thus enriching these comments. This gives officials as well as unbound helpers the opportunity to rate and filter messages, thus making these sources usable to support their work.

The aforementioned systems and approaches are used to make sense of huge data sets in a quantitative manner. Especially automated language processing is a complex, yet important task for these approaches that is prone to create e.g. false positives. For our analysis, however, we chose a qualitative approach by only analyzing a subset of user comments of both apps. We chose the following way to analyze user comments inspired by [30] and [32]:

1. Gathering user comments from Apple iTunes and Google Play Store from the apps KATWARN and NINA between 12/01/2015 and 02/09/2016. It turned out that reducing the comments to the actual version did only return a small number with less qualitative feedback, also leaving issues regarding e.g. faulty app versions aside. We then decided to consider all versions but to limit the period to generate a manageable and probably significant set of comments.

2. Manual classification [30] of the comments into topics based on semantics for each app. In case a comment contains more than one topic, it is being assigned to all relevant topics by its unique identifier. If equivalent topics could be identified in comments of both apps they were summarized. For our analysis the qualitative evidence of the messages is important, why findings of both app stores are summarized too. Only topics with more than one occurrence were taken into consideration to exclude single opinions.

3. Categorization of topics of both apps to make the findings comparable. The categories were built out of the set of topics, thus not being predetermined as [32] did. This allows unbiased categorization probably leading to a more accurate assignment.
The results may be subject to false classification as it was conducted based on the semantic of comments and are prone to the researchers’ biases. However, these results constitute the foundation to correlate it to the PSD model [16].

3.2 Persuasive System Design model

Oinas-Kukkonen and Harjumaa [16] present a model to analyze e.g. apps with regard to its potential towards behavior or attitude change. We here present a brief summary of this model and how we will correlate it to the aforementioned method.

To analyze or implement a system using the PSD model one has to consider the three phases of persuasive systems management. That is (1) understanding the key issues behind persuasive technology, (2) analyzing the persuasion context and (3) analyzing an existent or implementing a new system.

Key issues behind persuasive technologies.
The following key issues build the foundation for designing persuasive systems:

1. *IT is always on*, meaning that persuasion happens constantly and iteratively and that user’s goals may change over time.
2. *Commitment and consistency needed*, meaning that users have to commit to the task and that they want to reorganize cognitive dissonances once made visible.
3. *Direct and indirect route*, meaning strategies to either persuade the users by arguments or cues, having different effects on endurance or motivation.
4. *Incremental*, meaning behavior or attitude change has to be done in many small steps and motivate users to do them directly instead of postpone them to the future.
5. *Open*, meaning to avoid false content, reveal who the designers are and what they intend, always keeping the users’ voluntary attitude in mind.
6. *Unobtrusive*, meaning to find the right opportune moment to engage with the user and not disturbing her performing their primary task (see below).
7. *Useful and easy-to-use*, meaning the system should be created with regard to usability and user experience.

Analyzing the persuasion context.
The persuasion context is important to be thoroughly understood in order to design persuasive systems.

Firstly, the *intent* has to be determined, meaning to understand who the persuader is and what type of change is to be achieved (behavior or attitude).

Further the *event* has to be examined, more precise in which environment and problem domain the technology will be situated (use context), the users’ personality like interests, needs or goals (user context) and what kind of technology is being used (technology context).

Lastly the *strategy* has to be considered, meaning the content and timing of messages presented to the user, implying a *direct or indirect route* of persuasion.
Design of system qualities.
After these considerations Oinas-Kukkonen and Harjumaa [16] propose a taxonomy of design specifications for persuasive system design that could be used to address the above mentioned problems and biases. The 28 different design principals are grouped into four support categories, namely primary task support, dialogue support, social support and system credibility support.

3.3 Combining both methods
To classify the findings from our manual analysis of user comments within the PSD model, we conduct the following steps:

1. Identifying the persuasion context based on an analysis of both apps and the descriptions found on the respective websites. Unfortunately, the chosen comments did not give any valuable direct indications regarding the persuasion context, why the authors had to rely on publically available information.
2. Assign the topics to the design principals by each support category.

4 Analysis

4.1 Manual classification of comments about KATWARN and NINA
For KATWARN ten comments from Apple iTunes and 134 from Google Play Store were gathered on February 9th 2016 for the aforementioned period. The latest iOS version was 2.0.9 and the latest Android version was 2.0.14. For NINA 19 comments from Apple iTunes and 40 from Google Play Store were gathered at the same day. The latest iOS version was 1.1.3 and the latest Android version was 1.1.5.

Table 1 shows all topics manually extracted from the data set and categorized by the categories shown in Table 2. The authors generate these categories according to the topics. Although the authors at this time have no interest to analyze the findings in a quantitative manner, the occurrences of topics within the comments is presented to give an idea about its distribution. The total numbers per topic and category help to understand its importance and are intended to serve as a weight.

<table>
<thead>
<tr>
<th>Topic</th>
<th>C</th>
<th>K</th>
<th>N</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems after an update</td>
<td>C5</td>
<td>114</td>
<td>0</td>
<td>114</td>
</tr>
<tr>
<td>Irrelevant or false alerts due to lacking geographical reference (county, radius, current location)</td>
<td>C1</td>
<td>0</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>General Praise</td>
<td>C4</td>
<td>21</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>Other warnings (i.e. bomb disposals) (partly) not listed</td>
<td>C1</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Messages received too late</td>
<td>C1</td>
<td>8</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Request for customizable warning sounds</td>
<td>C6</td>
<td>7</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>
Location based warning only possible via GPS  
Weather or flood warnings are not displayed  
App crashes  
No warnings received  
Praise for well specified functionality  
Irrelevant or false alerts due to lacking geographical reference at night time annoy  
Confirmation that app works again after an update  
Mobile device specific problem  
Errors while navigating in the app  
Emergency advice functionality does not work  
Push messages do not work reliably  
High battery drain  
Requirement to add more control centers  
Test warnings annoy  
Incomplete warnings (i.e. all-clear without prior warning)  
Unclear privacy statement or permissions  
App not tested before rollout  
App loses credibility due to faulty updates  
Warnings to be protocollled and confirmed by users

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Category</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Quality of warnings</td>
<td>74</td>
</tr>
<tr>
<td>C2</td>
<td>Hardware or OS specific problems</td>
<td>3</td>
</tr>
<tr>
<td>C3</td>
<td>Problem with manufacturer</td>
<td>7</td>
</tr>
<tr>
<td>C4</td>
<td>Praise</td>
<td>38</td>
</tr>
<tr>
<td>C5</td>
<td>Malfunction of app</td>
<td>134</td>
</tr>
<tr>
<td>C6</td>
<td>Requirement</td>
<td>12</td>
</tr>
</tbody>
</table>

### 4.2 Embedding the found topics and categories into the PSD model

First the persuasion context of both apps is being described, based on findings in the apps itself as well as from the respective websites, as no evidence could be found in the users’ comments.

**Intent:** The persuader of both apps is the user himself, so the intent can be considered as autogenous, because the app is being installed by the user voluntarily. It can be assumed, that the user wants to be aware about warnings in his local area. However, it is the (local) authorities who may want to persuade the users to mind their behavior in certain threatening situations. Thus, also an endogenous intent can be seen. This also implies that the change type of both apps can be considered towards behavior change directly by the authorities or indirectly by reliable weather warning services, letting the user consider their behavior. There is no direct or indirect evidence the app is intended to change users’ attitude.
Event: The use context of both apps will also not be distinguished as the problem domain is identical. All German citizens constitute the target group and the persuasive system in both cases is a mobile smartphone app. Characteristics of the problem domain were mentioned in the introduction and therefore are not being repeated. As the app is usually not being distributed, but has to be installed voluntarily, users with an interest of mitigation and preparation towards emergencies might be the majority of the active users. This could enrich information regarding the user context. However, as both apps treat all users as a single audience, no more implications in this matter will be considered. Finally, the technology context of both mobile apps is also obviously the same. Due to their pervasive use, mobile devices have the potential to persuade users in-situ but on the other hand bare the risks of doing this ineffectively by annoying them or by technical constraints, like battery life or network coverage.

Strategy: As mentioned before, both apps focus on the direct route, as both send messages with clear suggestions of how to behave in the specific emergencies, i.e. keeping doors and windows closed during nearby major fires or chemical accidents. NINA additionally provides general behavioral information regarding different kinds of emergencies like storm, fire or flood in addition to warning messages. The timing of the messages is solely depending on the officials’ source systems. However, considering user comments especially regarding the quality and timing of messages, the entire category C1 (quality of warnings) can be applied here. Users of both apps complain about delayed, incomplete, inconsistent or irrelevant messages due to lacking geographical reference or disturbing messages at night time.

Design principals: Four topics were assigned to design principals of primary task support. The topic ‘praise for well specified functionality’ (K) is assigned to reduction, as it refers to appropriately reducing the users’ effort to get the desired information as a foundation of behavior change. The topics ‘requirement to add more control centers’ (K) and ‘warnings to be protocolled and confirmed by users’ (both apps) are assigned to tailoring, as both request for specific needs of certain user groups. The ‘request for customizable warning sounds’ (K) is assigned to personalization, as this setting aims at users’ individual preferences and situations, i.e. at work or at home. One topic is assigned to dialogue support, that is ‘general praise’ which refers to liking. Only the topic ‘unclear privacy statement or permissions’ can be assigned to trustworthiness of the category credibility support as proposed by [16]. This shows that users mistrust the designer’s intentions, probably because of unclear or missing explanations. However, more issues regarding credibility will be discussed in the following chapter. There is no topic found to be assigned to social support.

The topic ‘confirmation that app works again after update’ and all topics of the categories ‘malfunction of apps’ and ‘hardware or OS specific problems’ cannot be assigned to the PSD model and will be discussed in the following chapter too.

5 Discussion

The gathered data show that most users complain about the functionality and stability of both apps. One significant issue is that after an update of KATWARN in January
2016 the app crashed, leading to 114 complaints about it. This is by far the most frequent topic. Interestingly, after the issue was solved a few days later, a few users left a corresponding comment. Besides problems regarding the functionality of the apps, 74 comments deal with the quality of warnings being mainly either sent at inappropriate locations or times, if at all. However, there are also 38 comments praising the apps, mostly in general. There are also complaints regarding the manufacturers of both apps, like refused requirements, unclear privacy statements or untested updates. Lastly, some users complain about problems regarding the used hardware or operating system (OS). There are also twelve comments proposing new requirements.

Apart from the users’ comments, KATWARN only once took the opportunity to answer to a comment and thus to probably take influence. NINA support however answered several comments, interestingly only on January 4th and January 6th 2016 but retroactively for almost the whole review period.

The analysis shows that many of the found topics cannot be assigned to the persuasion context or the design principals of the PSD model. Topics regarding the malfunction of apps (C5) or hardware or OS specific problems (C2) show the significant influence this has on the trustworthiness or reliability of the system. Error-freeness is an implicit requirement with regard to the postulates of persuasive system design [16]. Hence malfunctions hinder systems to bare its persuasive potential. Although the support category dealing with credibility support addresses similar issues on a content level, faulty apps or device issues are located on the functionality level.

The many topics regarding the quality of the messages show that both apps have a deficit in using a proper strategy. Annoying the users with messages at inappropriate times or irrelevant places also reduce the possibility of persuading users to change their behavior, thus also contributing to loss of credibility.

There is a third basic problem regarding credibility. Some topics refer to problems regarding the manufacturers of the apps (C3) such as updates being released without prior testing.

Our findings confirm that these problems hinder both apps to develop their full persuasive potential. The fact that some users write comments after issues are solved by a new version (due to a faulty update) is however promising.

Another finding is that the user comments do not give significant evidence with regard to design principals proposed by the PSD model. Only the design principals reduction, tailoring, personalization and liking are assigned to individual topics. Interestingly, all these topics can be seen as positively connoted suggesting that these principals add to the users’ motivation to use the app. This finding, however, is not surprising as again both apps apparently were not designed with regard to persuasive technology but rather intend to inform the masses homogeneously, leaving behavioral changes to the users themselves by mostly only informing them.

With regard to our methodology, we conclude that analyzing user comments in app stores have some weaknesses to determine the persuasiveness of apps with regard to the PSD model. On the one hand this is due to apps are not always designed with regard to persuasive technology and on the other hand user comments in app stores apparently are mostly written when errors occur rather than focusing on content issues. However, we think that our methodological approach could be helpful to identi-
Warning as well as instructing affected citizens during emergencies is still challenging. Within our paper, we aim at examining mobile warning apps with regard to their persuasiveness. We therefore analyze user comments of two of the most important mobile warning apps in Germany with regard to persuasive system design. We have combined a qualitative mixed methods approach by manually extracting topics and categories of the chosen set of comments and then applying these to the persuasive system design model. The findings show that analyzing user comments with regard to persuasive system design for apps not intentionally aimed at changing users’ behavior, does not seem to be an effective method. Only very few topics or categories could be assigned to the PSD model. However, the main purpose of these apps is warning citizens by sending warning messages which both apps lack in quality and timing regarding to the findings. This highlights the importance of well-chosen messages when using the direct route. Another significant finding is that malfunctions of the apps lead to a high amount of user complaints which highlights the importance of error-freeness as an implicit requirement for systems to unfold its persuasive potential. The results also show that although the way of addressing the users by messages sent to smartphones and thus users individually, both apps neither aim at addressing users this way nor providing any functionality with regard to social support to enhance behavioral change.

As future work we will focus on the design of mobile warning apps by applying persuasive design methods such as the PSD model, as their intention obviously are to change users’ behavior based on warning messages. Especially the content, personalization as well as timing of these messages should be taken into special consideration, as these could be on the one hand the reason for proper preparation for or coping with emergencies or on the other hand lead to a tendency to ignore these warning messages. Another interesting aspect of developing mobile warning apps is how social support features could enhance citizens’ behavior, especially in the mitigation or preparedness phases of an emergency. Also collaborative development of target behavior amongst the users themselves could be of great interest, possibly revealing collective autogenous intentions and how to support its emergence.

The research project ‘KOKOS’ is funded by the German Federal Ministry for Education and Research (No. 13N13559).
8 References

2. Wachtendorf, T., Kendra, J.M.: Improvising Disaster in the City of Jazz: Organizational Response to Hurricane Katrina.