ORIGINAL RESEARCH

Spontaneous utilization: A classic grounded theory of utilizing ambient displays in professional, large-scale agile software development environments

Jan Schwarzer[a] and Susanne Draheim[a] and Kai von Luck[a] and Qi Wang[b] and Christos Grecos[c]

[a]Department of Computer Science, Hamburg University of Applied Sciences, Berliner Tor 7, 20099 Hamburg, Germany; [b]School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley Campus, Paisley PA1 2BE, Scotland, United Kingdom; [c]Department of Computer Science, Arkansas State University, PO Box 9, AR 72467, USA

ARTICLE HISTORY
Submitted: Mar 16, 2021; Revised: Jul 12 & Sep 17, 2021; Accepted: Nov 1, 2021

ABSTRACT
This research addresses an interdisciplinary problem area concerning the long-term deployment of socially embedded technology in authentic environments. It concentrates on the case of ambient displays, where long-term research in the real world is still scant and evidentially requires methodological development. This study seeks to advance on this situation at both findings and methodological levels. To this end, we introduce our Ambient Surfaces solution that was deployed in the agile software development department of a company for circa 5 years. Classic grounded theory was chosen to methodologically guide the evaluation, while the theoretical contribution of this work is a substantive theory exemplified by its core category of Spontaneous utilization. The theory reveals insights on how ambient displays are utilized by practitioners in professional and large-scale agile environments. We found, among others, that staff members used the Ambient Surfaces largely not on purpose, that our solution evolved towards having a strong emphasis on progress tracking information, and that inter-team awareness as well as intra-team communication were encouraged.

KEYWORDS
Ambient displays; field deployment research; grounded theory; large-scale agile software development; methodological development; substantive theory

1. Introduction

Investigating technology in situ fundamentally warrants new ways of how research is conceptualized as people and practices are more than just their relationships with technology (Bjørn & Boulus-Rødje, 2015). There exists a symbiosis of humans with smart environments that goes well beyond technical boundaries (Stephanidis et al.,

This is an Accepted Manuscript of an article published by Taylor & Francis in the International Journal of Human–Computer Interaction on 20 Dec 2021, available online: http://www.tandfonline.com/10.1080/10447318.2021.2002051. Contact: jan.schwarzer@haw-hamburg.de.

2019). Practices and technology become intertwined and embody continuously changing entities that are redesigned and reorganized (Bjørn & Boulus-Rødje, 2015). In-situ evaluations have gained momentum throughout the domain of human–computer interaction (HCI), but particularly in the disciplines of computer-supported cooperative work (CSCW) and ubiquitous computing (Ubicomp) (Siek, Hayes, Newman, & Tang, 2014). While in-situ research is considered resource-intensive, it simultaneously expands on authentic usage and highlights how technology interacts with the environment (Siek et al., 2014). To be sure, the complexities of operational feasibility can only be determined in the field (Nunamaker, Briggs, Derrick, & Schwabe, 2015), where social aspects are becoming a constitutive part of the research (Bjørn & Boulus-Rødje, 2015). Contrary to other types of research, in-situ research enables the investigation of long-term effects and provides researchers with a means to cope with issues that arise from a technology's novelty (Alt, Schneegaß, Schmidt, Müller, & Memarovic, 2012; Ojala et al., 2012). Consequently, the ecological validity of data obtained from in-situ research is assumed high (Alt et al., 2012). Despite its challenging nature (Jurmu et al., 2016), in-the-wild research affects both science and society most notably (Nunamaker et al., 2015). Recent contributions from a variety of disciplines such as information systems (e.g. Nunamaker et al., 2015), HCI (e.g. Jurmu et al., 2016), information visualization (e.g. Preim, Ropinski, & Isenberg, 2018), CSCW (e.g. Bjørn & Boulus-Rødje, 2015), and Ubicomp (e.g. Hazlewood, Stolterman, & Connelly, 2011) stress the relevance of such endeavors. Some studies specifically emphasize on calls for more longitudinal in-the-wild research (e.g. Preim et al., 2018) as it allows us to scrutinize how a technology is adapted in the real world (Preim et al., 2018; Siek et al., 2014).

This research draws attention to the field of ambient displays—a sub-area of the Ubicomp discipline (Mankoff et al., 2003). Studying pervasive technology in situ has gained notable interest (Keskinen et al., 2013). However, research on ambient displays rarely attempts to scrutinize them in the wild over a longer period of time and concentrates primarily on short-term deployments (Du, Degbelo, & Kray, 2017; Mäkelä, Heimonen, & Turunen, 2018). More empirical knowledge is required to, for instance, determine both their actual value to users (Parker, Tomitsch, Davies, Valkanova, & Kay, 2020) and their potential to foster collaboration (Ardito, Buono, Costabile, & Desolda, 2015). Fundamentally, long-term investigations of ambient displays in the wild require methodological development that guides researchers in such endeavors (Jurmu et al., 2016; Mäkelä et al., 2018). There is a lack of standard methods for evaluating ambient displays regarding, for instance, their effectiveness and usability (Shelton & Nesbitt, 2017). Research is desired that brings forth contributions on a methodological, theoretical, artifact, opinion, and survey level (Du et al., 2017).

In response, this paper elaborates on methodological and empirical insights from our longitudinal in-the-wild study. An investigation of our custom ambient display solution—henceforth referred to as *Ambient Surfaces*—is presented. The solution embodies a special sub-class of ambient displays that leverages screen-based solutions and is concerned with "supporting informal, nonurgent communication, collaboration, and awareness" (Huang, Mynatt, Russell, & Sue, 2006, p. 37). For circa 5 years, our solution was deployed in a professional, large-scale agile software development (ASD) environment. Evidently, investigations of ambient displays in the software development domain are, among others, often short-termed and miss rigorous evaluation techniques (Bedu, Tinh, & Petrillo, 2019). While agile teams are still in need for tools supporting collaborative team practices, the literature underlines the value of utilizing large displays for this task (Scott-Hill et al., 2020). Classic grounded theory (GT) was chosen to constitute the methodological foundations for this research. Due to the methodology's premise of, for instance, striving towards substantive or formal generality (Glaser, 1998), we argue that classic GT methodology encompasses promising means to tackle contemporary issues in HCI research such as scaling knowledge contributions across situations and contexts (Brown, Bødker, & Höök, 2017). Ultimately, this study's main, theoretical contribution is a revelatory substantive theory—exemplified by its core category of *Spontaneous utilization*—that stresses the very nature of the long-term usage by placing spontaneity at its center. With the word *utilization* we refer to situations in which people engaged actively or passively with the Ambient Surfaces.

The remainder of this article is organized as follows: Section 2 introduces the related literature, while Section 3 illustrates both the research site and the Ambient Surfaces solution. Subsequently, Section 4 elaborates on epistemological and methodological decisions as well as data collection and analysis procedures. Section 5 then presents the results of this study. Following this, Section 6 discusses the theoretical contribution and its implications, research limitations, and recommendations for future research. Finally, Section 7 concludes this article.

2. Related work

Initially, a background on ambient displays is provided. Afterwards, methodological advances concerning ambient displays research are presented. Finally, we concentrate on the application of ambient displays in the ASD context.

2.1. Background on ambient displays

While the first known ambient display dates back to 1995—the *Dangling String* (Weiser & Brown, 1995)—, there is no uniformly accepted definition (Shelton & Nesbitt, 2016). Ambient displays can be considered "aesthetically pleasing displays of information which sit on the periphery of a user's attention" (Mankoff et al., 2003, p. 169). They generally show the features of what Weiser and Brown (1995) termed *calm technology* in the 1990s. More specifically, ambient displays are characterized by the following attributes (Pousman & Stasko, 2006): they display important information; they can move from the periphery to a person's attention and vice versa; they are a physical representations (i.e. tangible objects or screens); they indicate subtle changes of information; and they are environmentally appropriate as well as aesthetically pleasing.

Following Börner, Kalz, and Specht (2013), ambient displays find application in personal, semi-public, or public contextual levels, while their individual characterizations are diverse and multifaceted. They address the different forms of sensory perception including vision, hearing, haptic, odor, and taste. Besides their core characteristics of distributing non-critical information and establishing informational awareness, their aesthetic features and the decorativeness play an important role as well. The user experience of ambient displays radically stands in contrast to traditional task-orientated situations (Hazlewood et al., 2011). For instance, they attract people standing nearby to engage in interactions (Ardito et al., 2015). Ambient displays are meant for ad-hoc and serendipitous use (Ghare, Pafla, Wong, Wallace, & Scott, 2018). Ghare et al. (2018) as well as Elhart, Mikusz, Mora, Langheinrich, and Davies (2017) urge us to consider the wider context surrounding ambient displays, while Williamson and Williamson (2014) emphasize the existing reciprocal relationship between technology and space. This may give reason as to why ambient displays need to be understood in the wild and not in lab-based environments (Ardito et al., 2015).

Usually, research on ambient displays is short-termed (Du et al., 2017; Mäkelä et al., 2018) and studies urge to, for instance, investigating the aspect of user acceptance in real life for longer periods of time (Colley, Häkkilä, Forsman, Pfleging, & Alt, 2018). To the best of our knowledge, the work from Ojala et al. (2012) is one recognized exception to this rule—the authors deployed their *UBI hotspot* solution in the public space for several years. Consequently, it is still one of the most important issues to find useful adoption scenarios for ambient displays in real-world environments (Koch, Ott, & Richter, 2014) and, simultaneously, to construct still lacking general theories on these devices (Alt et al., 2012). It remains a challenge to address factors such as a display's location and the relevance of content as regard to user engagement (Parker, Tomitsch, & Kay, 2018). Existing knowledge is limited regarding how and when ambient displays provide value to their users (Parker et al., 2020). In addition, we know little about how these devices can potentially foster collaboration (Ardito et al., 2015). Fundamentally, the design of ubiquitous applications remains a challenging task (Marquardt & Greenberg, 2015) and guidance is warranted for those who plan to deploy such systems to engage users (Ghare et al., 2018).

2.2. Methodological advances in ambient display research

Throughout the formative years (i.e. 1997–2003), ambient displays were typically not evaluated at all (Mankoff et al., 2003). Subsequently, scrutinizing pervasive technology in the field gained attention (Keskinen et al., 2013), while it simultaneously increased associated difficulties. Hazlewood et al. (2011, p. 877) describe this vividly in saying that researchers were now facing the situation of providing "in-depth evaluations on something that is defined as blending with the surrounding world, and meant to be (in some respects) ignored." Researchers started to embark on investigating ambient displays in the context of pilot studies (i.e. mostly laboratory experiments and methods) and in the form of in-depth field studies (Börner et al., 2013). While pilot studies lasted at most several weeks, field studies went from several days or weeks to long-term deployments lasting several months. Utilized data collection methods were interviews, observations, surveys, focus groups, and interaction logging (Alt et al., 2012). Typically, however, researchers often found themselves in needing to create own approaches to evaluate their individual solutions (Keskinen et al., 2013).

In the past decade, the literature started to, on the one hand, acknowledge the difficulties associated with evaluations of ambient displays, while also calling for more long-term in-the-wild studies (Börner et al., 2013). On the other hand, research began criticizing existing evaluations because they were predominately conducted in laboratory environments (Ojala et al., 2012). While there is a trend towards scrutinizing ambient displays in contexts such as cities, universities, and schools, Ardito et al. (2015) still found that as much as 43% of the 206 included studies in their literature survey were conducted in laboratories. Evidently, the evaluation of ambient displays in the wild remains challenging (Mäkelä et al., 2018; Williamson & Williamson, 2017). In a recent systematic review including 459 studies from 1996 to 2016, Shelton (2020) found that as much as 70 % (320) of them conducted evaluations to some extent. While most research was conducted using only qualitative methods (n=166), many studies also leveraged a mixed-methods approach (n=131). Only a few studies emphasized entirely on quantitative methods (n=20) or were not clear in their documentation (n=3). A fundamental tendency of using discrete data collection techniques as opposed to holistic methodological approaches can be observed. We identified surveys, heuristic inquiries, and experimental research to be the only methodologies mentioned.

Unsurprisingly, while there is a steadily increasing number of evaluations observable (Shelton, Nesbitt, Thorpe, & Eidels, 2020), there is simultaneously a need for methodological contributions (Du et al., 2017). Specifically, recent studies stress existing epistemological (Jurmu et al., 2016) and methodological challenges (Schwarzer, von Luck, Draheim, & Koch, 2019) and, accordingly, argue for both methodological development (Du et al., 2017) as well as improved evaluation methods for long-term in-situ studies (Mäkelä et al., 2018). There is a need for devising new standard methods for evaluations (Shelton & Nesbitt, 2017) which requires studies emphasizing on long-term field deployments to elicit the very nature of ambient displays—"to become truly *ambient* within a given environment" (Hazlewood et al., 2011, p. 879, emphasis as in original). Methodologically, recent studies introduce the ideas of, for example, increasing automation throughout the analysis process (Mäkelä et al., 2018) or incorporating data collection methods such as virtual reality tools (Mäkelä et al., 2020). However, limitations remain with regard to studying complex behavior (Mäkelä et al., 2018), the level of ecological validity compared to real-world deployments (Mäkelä et al., 2020), and the ability to study long-term or continuous use (Mäkelä et al., 2020). Ultimately, Shelton et al. (2020) urge researchers to, epistemologically and methodologically, deal with the fact that for full evaluations, future studies must be conducted in real-world environments where participants are familiar with the conveyed information.

2.3. Ambient displays in agile software development

Considering ambient displays in the ASD context calls attention to the field of software visualization. Software visualization is a sub-area of information visualization that aims to foster understanding and insight in the software engineering process by visualizing the structure, behavior, and evolution of software (Diel, 2007). According to Paredes, Anslow, and Maurer (2014), ASD teams utilize a variety of visualization tools (e.g. dashboards, information radiators, and ambient displays) for designing, developing, communicating, and progress tracking purposes. These tools assist in facilitating knowledge sharing and awareness among team members in practices such as code reviews or software maintenance (Paredes et al., 2014).

Large displays are a common means to convey information in shared software development workspaces (Ye, Ye, & Liu, 2018). For example, Biehl, Czerwinski, Smith, and Robertson (2007) and Jakobsen et al. (2009) use teamwork awareness tools in the form of large wall projections. *FASTDash* (Biehl et al., 2007), a repeatedly cited tool, concentrates on immediate awareness of conflict situations throughout the software development process. In contrast, *WIPDash* (Jakobsen et al., 2009) indicates the project status by providing information from a team's software repository. Also *TeamWATCH* (Ye et al., 2018) utilizes large displays to assist software developers in maintaining group awareness and fostering collaboration. In a similar vein, *DashVis* (Scott-Hill et al., 2020) supports meetings by displaying progress tracking visualizations on large touch screens. Another example is *CodePad* (Parnin, Görg, & Rugaber, 2010), which, in contrast, uses additional displays in a software developer's personal workspace to, among others, assist in maintaining awareness of development artifacts.

The application of visualization tools in the ASD domain is, however, limited. Most of them are solely build for single-user scenarios as to they are bound to desktops and integrated software development environments (Scott-Hill et al., 2020; Sharma, Mehra, Kaulgud, & Podder, 2019). In doing so, such tools hamper working collaboratively

(Scott-Hill et al., 2020) and continue the prevalent trend to not leverage the available physical space to, for instance, create aiding information radiators (Sharma et al., 2019). This issue, arguably, becomes particularly profound in the context of large-scale (i.e. 2–9 agile teams) and very large-scale (i.e. $\geq$ 10 agile teams) ASD environments, where knowledge sharing and inter-team coordination are still challenging (Dingsøyr, Moe, Fægri, & Seim, 2018). Agile methods in these contexts are still contested and only a few empirical studies are available (Rolland, 2016). Existing tools further largely concentrate on one single source of information, mostly focus on a developer's real-time information, and typically miss to provide, for example, historical information in the form of statistical results (Ye et al., 2018). Additionally, long-term studies of ambient displays in professional ASD contexts are rare (Schwarzer et al., 2016).

Secondary studies in the software visualization domain indicate a lack of real-world examples and rigor in evaluations. For instance, Paredes et al. (2014) found that the main issue is the little research conducted on how to evaluate software visualization tools. There is a gap between what is being studied in the scientific community and what is being adopted in the real world. In a systematic review on the topic of software visualization evaluation, Merino, Ghafari, Anslow, and Nierstrasz (2018) conclude that 62% of the 181 investigated studies show no evaluation at all or solely anecdotal evidence. Reportedly, this may explain the low adoption of software visualization tools. A recent tertiary study (Bedu et al., 2019), including both aforesaid studies, states the following conclusions regarding software visualization: firstly, evaluations of software visualization tools remain largely superficial, are often short-termed, and lack rigorous techniques; secondly, there is a crucial need for empirical studies; and, finally, human factors in evaluations of software visualization tools require attention.

3. Research site and custom ambient display solution

Below, we introduce both the research site and the Ambient Surfaces solution.

3.1. Werum IT Solutions GmbH

A German company named *Werum IT Solutions GmbH*[1] (hereafter referred to as *Werum*)—specifically, its ASD department—participated in this study. The company specializes in developing manufacturing execution systems and IT solutions for the pharmaceutical and biopharmaceutical industries. Roughly 70 to 80 people were employed in the ASD department. Based on an online survey (n = 35), almost 90% of them were older than 30 years old, while the majority (48.6%) ranged between 31 to 40 years of age. Circa 80% of the staff were male. Most employees were software developers (62.9%), followed by Scrum Masters and Product Owners (both 14.3%), team leaders, heads of department, administrators, and others (each 8.6%). Roughly 77% of staff members had been working at Werum for at least 3 years.

Werum initiated a transition to the agile method *Scrum* (Schwaber & Beedle, 2001) roughly a year prior to the present research in early 2013. Hence, the company could be considered matured in applying agile practices at the point of commencing the field deployment in February 2014 (Dybå & Dingsøyr, 2008). The ASD department, including its changing number of agile teams in the course of this research (i.e. four to eight teams in total), can be characterized as a large-scale ASD environment (Dingsøyr,

[1] `https://www.werum.com/de/home/`

Figure 1. The setup in the newly constructed two-story building for the ASD department staff (2017). In the bottom-right corner, there is a floor plan from the ground level indicating the location of the displays (blue dot). Abbreviations: O=Office; T=Toilets; CR=Conference Room; P=Printers; K=Kitchen; S=Storage Room.

Fægri, & Itkonen, 2014), while the ASD teams work in a co-located workplace.

The staff is equipped with a variety of tools that assist them in developing the company's software product. For the context of this research, the following tools deserve a special mention: *Atlassian Jira* (e.g. to store user stories), *Atlassian Confluence* (e.g. to share architectural decisions), *Jenkins* (for continuous integration purposes), *GoCD* (to automate the build and deployment infrastructure), *Tetris* (a custom tool to display test summaries), and *Avatar* (a custom tool to show graphical test metrics).

3.2. Ambient Surfaces

The Ambient Surfaces solution (see Figure 1) consisted of three principal components: the custom software application, interactive displays, and compact desktop computers. The software application was responsible for handling touch interactions with the displays, selecting data from the aforesaid tools, and preparing the corresponding visualizations. It built on *Windows Presentation Foundation*[2] framework to render user interfaces. This software suite was utilized because the authors already had experience with it from previous studies. In total, two displays were utilized throughout the study (each $\geq$ 46 inches in size). The first display was deployed in February 2014, the second one followed on request in August 2015 because staff members perceived the available space on one display insufficient to visualize all the different content. Both displays were mounted on a rack with rolling wheels in a landscape configuration. The total height of each installation was roughly 1.80 meters. Both monitors provided a 1080p resolution (i.e. 1,920 $\times$ 1,080 pixels) and infrared touch sensors allowing the simultaneous detection of between 2 and 32 touches respectively. Two compact desktop computers were utilized to operate the displays and to run the software application.

[2] https://docs.microsoft.com/en-us/dotnet/framework/wpf/

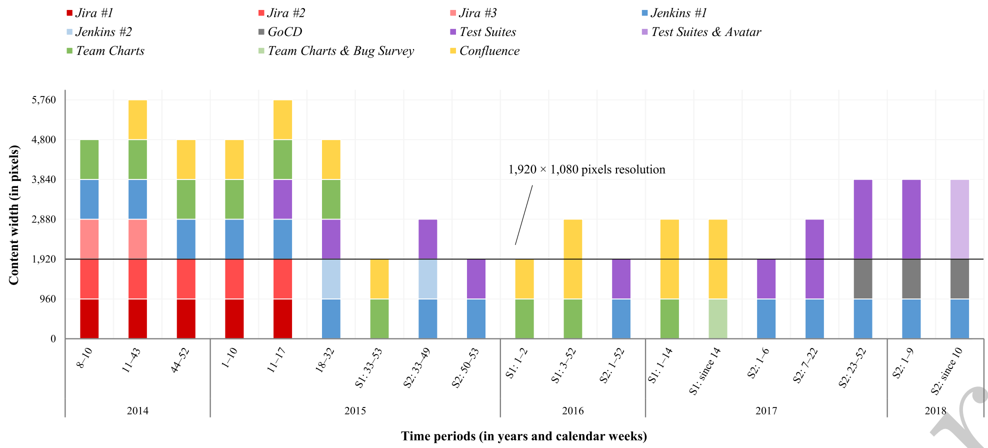

Figure 2. A bar chart diagram based on Schwarzer et al. (2019) showing the evolution of both Ambient Surfaces (i.e. system *S1* and system *S2* as of Week 33 in 2015) over time by relating added, removed, and resized (i.e. increases in width) information views to actual display widths (i.e. in pixels) and time periods.

Both machines were only allowed to access resources in the company's intranet. There were two distinct common areas in which both Ambient Surfaces were situated due to the construction of a new building for the staff in 2017. Both common areas were selected because they seemed to provide the chance of opportunistic interactions (Parker et al., 2020). These areas had similar characteristics such as storing beverages, providing a pool table (first location) or a table soccer installation (second setup), and showed a notable number of passers-by. The biggest difference was that prior to the relocation in 2017, solely the ASD department staff had direct access to both systems. This changed in the new setup as the building incorporates a canteen which is used by the entire staff during opening hours (i.e. 11:30–14:00).

Eight information views were deployed throughout the study (see Figure 2), including the views labeled *Jira* (i.e. Sprint activities), *Jenkins* (i.e. build statuses), *Team Charts* (i.e. team charts), *Confluence* (i.e. news), *Test Suites* (i.e. test suite statuses), *GoCD* (i.e. pipelines and stages statuses), *Bug Survey* (i.e. bug charts), and *Avatar* (i.e. test suite summaries). An example is shown in Figure 3 that illustrates the *Avatar* view, which displayed charts that were incorporated as a static screenshot. Depending on the data, the software application was configured to display either two views at a time or one view on its own. The views utilized different means to present data, whereas some were entirely custom-made. Data sources were incorporated via APIs provided by the tools or they were included as is (e.g. a Wiki website). Changes to this set of information views were motivated by, for instance, views showing too much information (e.g. the *Jira* views were removed in 2015), views hiding information due to design-related issues (e.g. the *Confluence* view's width in pixels was increased in 2016 on system *S1*), and views becoming obsolete (e.g. one *Jenkins* view was removed in 2015 on system *S2*). In most cases, Werum employees proactively contacted the research team to ask for one of such revisions.

The design process was guided by, as Parker et al. (2020) suggest, the value that our solution may deliver to staff members. For example, the *Confluence* view allowed employees to contribute own content hence create ownership and control about messages being spread (Parker et al., 2020). Also, the interaction modalities were kept to a minimum (i.e. scrolling gestures and selection) to avoid frustration of users (Ardito et al., 2015). With data stemming from different intranet sources, we additionally

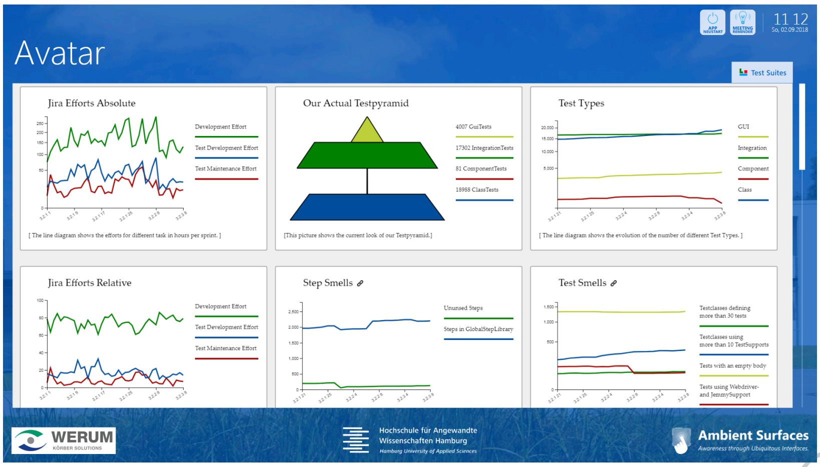

Figure 3. The *Avatar* view as of 2018. The clock in the right upper corner was configured to display abbreviations of weekdays in German (i.e. "So" corresponds to Sunday). The German label in button "App Neustart" refers to the process of restarting the software application (i.e. analogously "App Restart").

aimed at providing a rich information offer. It was geared towards preventing from issues (e.g. incomplete and incorrect information) seen with awareness tools that yield information from only one data source (Ye et al., 2018). Simultaneously, the design built on previous, preliminary research back in 2012 and 2013 where we deployed the solution in two other professional agile contexts. During this time, we gathered first experiences in terms of, for instance, reducing certain information to a minimum (e.g. with respect to displaying team activities) or recoloring information views (e.g. in the case of broken builds). Normally, we suggested initial design drafts which were then revised according to feedback, while employees at Werum also came up with own ideas for new information views that we then collaboratively realized (e.g. the *Avatar* view).

4. Methodology

The epistemological and methodological foundations as well as data collection and analysis procedures are now introduced.

4.1. Classic grounded theory

The existing methodological challenges of ambient display research required us to approach our enterprise rather broadly. Our research set out with no predefined research questions or hypotheses. Instead, we were motivated by a general curiosity that guided us in the early stages of the research. We concluded that the philosophical stance of pragmatism fitted our research the best. We were convinced that both qualitative and quantitative methods could be potentially of use. Similarly, we believed we needed to be objective and subjective in our epistemological orientation. Pragmatism's notion of seeing research as a continuum between those two opposing poles seemed most promising to us. As we did not define any hypotheses in advance, we did not commence our research with a deductive stance. Similarly, our initial position was not characterized by determining how well deduced consequences of a hypothesis correspond to facts (i.e. induction). Rather, we adapted an abductive lens "to infer the best possible

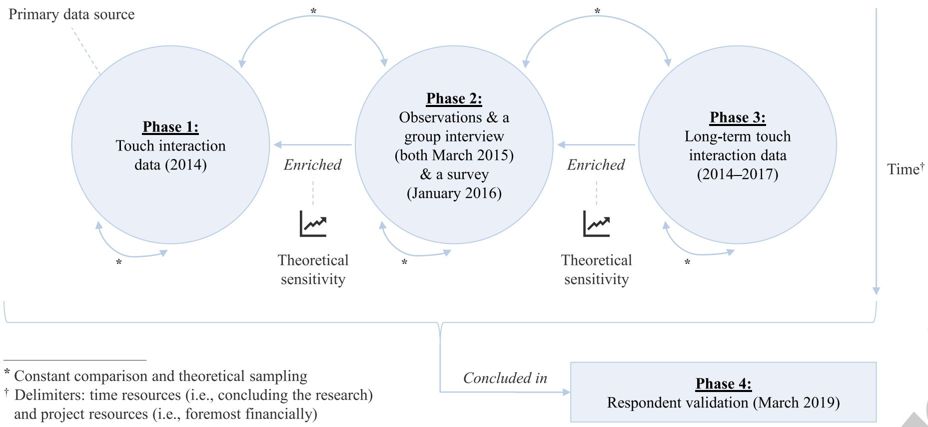

Figure 4. The four research phases and the different types of data utilized.

explanation from the data in whatever form it is presented" (Bryant, 2009, p. 30).

Ultimately, we decided for classic GT methodology (Glaser, 1978, 1992; Glaser & Strauss, 1967). With that decision, we went against creating a new methodological approach which is often seen, for example, in software visualization studies (Mattila et al., 2016). Classic GT provides a means to think about and theorize from data—a conceptualization process that results in a theory about a substantive area (Morse et al., 2009). It seeks to identify and to elaborate on a central concept (i.e. the *core category*) that emerges from the data by constant comparison. The core category is, fundamentally, the overriding patterns found in the data (Glaser, 1998). Our motivation to choose classic GT was threefold: Firstly, classic GT is compatible with a pragmatist perspective. Both have in common their reliance on observation and insight, while being equipped with never-ending efforts to comprehend, persuade, and enhance (Bryant, 2009). The methodology shows also a firm association with abductive inquiry due to its ability to allow thinking of data theoretically (Bryant, 2009). Secondly, classic GT provides flexibility. It embodies a qualitative analysis approach inviting any kind of analysis that produces findings, qualitatively or quantitatively arrived at, not by statistical methods (Glaser, 1992). Thirdly, classic GT does not prescribe to define any research question or hypothesis in advance (Glaser, 2008).

4.2. Data collection and analysis

Practice-based research indicates that the emergence of practices demands time (Jurmu et al., 2016). Also display research phenomena such as the *novelty effect* (Koch, von Luck, Schwarzer, & Draheim, 2018) and *display blindness* (Müller et al., 2009) require attention. Fundamentally, we aimed at deploying our solution for the course of an entire PhD project. However, whether staff members kept utilizing the Ambient Surfaces, was a somewhat open question at the beginning. We were aware that technology falling into disuse is an issue discussed in the related literature (Huang et al., 2006). In the end, we were able to collect data over the course of circa 5 years (i.e. between February 2014 and March 2019). During these years, the research team kept in close contact to Werum staff members, particularly Scrum Masters via phone and email. We also regularly visited Werum on-site. In doing so, we were able to early identify technical problems and the need for revisions (e.g. to deploy a second screen).

The research process organization built on other GT studies that had structured

Figure 5. A memo wall in one of our offices back in 2014, showing different data visualizations such as pie chart diagrams and heat maps as well as user interface screenshots.

their research in different phases (e.g. Walsh, 2015). The present study happened to be organized in four phases, accordingly labeled *Phase 1* to *Phase 4* (see Figure 4). *Phase 1* only considered quantitative touch interaction data. Subsequently, a combination of both kinds of data sources in the form of observations, a group interview, and an online survey enriched the theory-generating process in *Phase 2*. *Phase 3* solely built on long-term quantitative touch interaction data. While in *Phase 1* and *Phase 2* data was progressively being analyzed, in *Phase 3* descriptive statistics were elaborated on retrospectively. *Phase 4* concluded this research by including feedback from a respondent validation to increase the authenticity of findings. It is noted that the four phases were not connected to or a result of any revision of the Ambient Surfaces. Pictorial models were utilized to conceptually and theoretically conclude each phase (Glaser, 1978) and to guide the process of writing up the theory as it was formulated. We made the touch interaction data, questions asked, and the observation template publicly available (Schwarzer, Draheim, von Luck, Wang, & Grecos, 2021).

Finally, this research builds on a step-by-step systematization of GT's constant comparison process that suggests four distinct criteria to be elaborated on throughout each comparison (Boeije, 2002): the aim of comparisons, the important questions asked, a description of activities, and the results (see Section 5). At the different stages, we leveraged the idea of *memos* as a basis for comparisons. Memos in GT methodology are a medium that do not underlie strict guidelines other than to informally capture ideas for the emergent theory as they occur (Glaser, 1998). We used memos in the form of, for example, printed tabular overviews or charts (see Figure 5).

4.2.1. Phase 1

Data collection We set out with guidance provided by Muller (2014), who encourages researchers to choose methods which allow them best to perceive and know. It was decided to commence this research with the Ambient Surfaces' custom touch interaction logging mechanism. This mechanism allowed touch interactions with the Ambient Surfaces' screen to be automatically recorded and happened to be used as the primary data source throughout the research. Selecting a primary data source is recommended in the literature (Stol, Ralph, & Fitzgerald, 2016), while our motivation was fivefold:

(1) Sensor data to track user activities is commonly utilized in ambient display research (Börner et al., 2013) and logging is considered helpful in long-term research endeavors (Alt et al., 2012).
(2) As it is crucial in ambient display research to collect data unobtrusively (Börner et al., 2013), the logging mechanism arguably allowed data to be collected without distracting users.
(3) Another reason was that phenomena such as the novelty effect necessitated attention during analyses. Some studies, for instance, reported to have extended their research due to this effect (e.g. Gallacher et al., 2015). As only a small number of the staff was involved in the preparation stage prior to February 2014, most of the employees were unfamiliar with the Ambient Surfaces at the time of the deployment. Hence, we expected increased interactions in the initial weeks due to reasons of novelty. We also foresaw such behavior appearing when changes were applied to the Ambient Surfaces. With the help of the logging mechanism, we aimed at approximating the duration of the novelty effect.
(4) Additionally, the logging mechanism kept the resources in this early stage of the research in check. We argue that methods such as observations and, for instance, interviews would have produced labor-intensive piles of data that would have shown little added value due to the prevalence of the novelty effect. Data collection techniques such as observations are known to account for time-intensive workloads (Corbin & Strauss, 2015; Mäkelä et al., 2018). However, we intended to use these methods later throughout the research.
(5) Finally, a huge portion of the logging mechanism was already implemented prior to this research hence it was readily available when we commenced this study.

Overall, there were two log files created by the logging mechanism: firstly, a file that stored *touch events* and, secondly, a file that included *view events*. Touch events were generic events in the software framework that were automatically triggered and caught by our solution when a person touched a screen's surface. In contrast, view events were individually programmed events that corresponded to the repertoire of actions the Ambient Surfaces' user interfaces provided (e.g. scrolling gestures to read articles). Compared to touch events, view events were lower in numbers because not every area of the user interfaces provided pre-defined functionalities, while touch events were triggered with every touch gesture. This larger number of events was also the primary reason to emphasize analyses of the novelty effect on touch events in one of our previous studies (Koch et al., 2018). The logging mechanism was revised following changes to account for, for example, view events of newly added information views.

For the purpose of generating theory, we largely concentrated on touch event analyses at the beginning, while we logged all event types continually throughout our research. In the touch event log file, three types of events were logged: *touch move* events (i.e. movements on the screen's surface), *touch down* events (i.e. placing one to many fingers onto the display's surface), and *touch up* events (i.e. lifting one to many fingers from the display's surface). Each interaction of a finger resulted in one touch down event, one to many touch move events, and one touch up event. For each of these events, a timestamp, its type (i.e. touch move, touch up, touch down), and a corresponding x and y coordinate were recorded.

Data analysis The aim of comparing touch interaction data was to unveil somewhat saturated latent patterns—or what Glaser (1978, p. 40) refers to as "directing hypotheses"—to guide future data collection activities. Emerging questions were doc-

Times of day	Total	Total (in %)
07:00	0	0.00%
07:30	40	2.03%
08:00	3	0.15%
08:30	64	3.24%
09:00	111	5.63%
09:30	50	2.53%
10:00	83	4.21%
10:30	133	6.74%
11:00	69	3.50%
11:30	21	1.06%
12:00	107	5.42%
12:30	128	6.49%
13:00	45	2.28%
13:30	**214**	**10.85%**
14:00	99	5.02%
14:30	89	4.51%
15:00	157	7.96%
15:30	107	5.42%
16:00	143	7.25%
16:30	34	1.72%
17:00	125	6.34%
17:30	90	4.56%
18:00	20	1.01%
18:30	30	1.52%
19:00	6	0.30%
19:30	3	0.15%
20:00	2	0.10%
Total:	**1,973**	**100.00%**

Weeks 21–30

Times of day	Total	Total (in %)
07:00	0	0.00%
07:30	18	1.38%
08:00	13	0.99%
08:30	34	2.60%
09:00	35	2.68%
09:30	62	4.74%
10:00	39	2.98%
10:30	28	2.14%
11:00	63	4.82%
11:30	13	0.99%
12:00	75	5.74%
12:30	**166**	**12.70%**
13:00	81	6.20%
13:30	114	8.72%
14:00	30	2.30%
14:30	100	7.65%
15:00	55	4.21%
15:30	11	0.84%
16:00	90	6.89%
16:30	128	9.79%
17:00	12	0.92%
17:30	48	3.67%
18:00	57	4.36%
18:30	2	0.15%
19:00	26	1.99%
19:30	7	0.54%
20:00	0	0.00%
Total:	**1,307**	**100.00%**

Weeks 31–40

Times of day	Total	Total (in %)
07:00	7	0.38%
07:30	9	0.49%
08:00	25	1.37%
08:30	44	2.41%
09:00	55	3.01%
09:30	42	2.30%
10:00	158	8.65%
10:30	91	4.98%
11:00	41	2.25%
11:30	142	7.78%
12:00	103	5.64%
12:30	**326**	**17.85%**
13:00	129	7.06%
13:30	35	1.92%
14:00	65	3.56%
14:30	101	5.53%
15:00	89	4.87%
15:30	46	2.52%
16:00	32	1.75%
16:30	134	7.34%
17:00	101	5.53%
17:30	23	1.26%
18:00	16	0.88%
18:30	0	0.00%
19:00	2	0.11%
19:30	8	0.44%
20:00	2	0.11%
Total:	**1,826**	**100.00%**

Weeks 41–50

Figure 6. Three instances of the crude index *Utilization of the Ambient Surface*, based on touch down event data from weeks 21 to 30 (left), weeks 31 to 40 (center), and weeks 41 to 50 (right) in 2014. Times of day with the highest number of utilizations are emphasized with a bold font and the darkest blue tone (the less the blue tone the less the number of touch down events).

umented and a stack of data created that could later be triangulated with other data sources. We did not target identifying the core category in this early stage of research. *Phase 1* was understood as a first step towards the goal of generating theory.

In early 2014, we focused the analysis on two themes. Firstly, we wished to identify times of day, in which people used the first Ambient Surface most notably. This choice was somewhat random as to we could have also emphasized on content utilization. We trusted in GT methodology's premise that categories and their properties emerge independently of where the research commences (Glaser, 1998). Secondly, we expected the novelty effect having an impact on the logged touch interaction data for some time. We wanted to know what data could be linked to this effect to cope with it during analyses hence better learn about baseline usage. Consequently, the following two questions were posed in advance for *Phase 1*:

(1) During what times of the day is the first Ambient Surface most prominently being utilized?
(2) How long can the novelty effect be notably observed in the data collected?

In *Phase 1*, the idea of crude indices was leveraged for analyses (Schwarzer et al., 2019). Principally, we used crude indices to quantify utilization in terms of mapping aggregated numbers of interaction to discrete time windows. To this end, a crude index named *Utilization of the Ambient Surface* was created using spreadsheet software (see Figure 6). This crude index built on a variable called *Timestamp of event* from the touch interaction log file (i.e. the timestamp of an event in milliseconds), while we focused the analysis on touch down events. We did so because we were not interested in the time span of an interaction (touch move events), but for the point in time of an interaction. For this purpose, both touch down and touch up events provided almost the exact same information regarding their timestamps. Principally, the variable *Timestamp of event* enabled to cast light on temporal questions regarding interactions and was thus selected. The created crude index contained two items: firstly, the item *Number of interactions*, which summed data from *Timestamp of event*; and secondly, the item *Times of day* that allowed *Number of interactions* to be organized in individual time windows. The working day was split into 27 half-hourly parts (i.e. from 7 am to 8 pm) to account for most interactions.

Analyses and log file downloads were scheduled on the weekends to avoid interfering with the use of the system during the daytime. As employees became more accustomed to the system over time, it was expected to see patterns observed in newly collected data becoming incrementally more similar. We noticed this development at the end of 2014. For example, touch interaction data from different time periods respectively indicated only one instance of very strong utilization (see the two-digit percentage values in Figure 6). As we elaborated elsewhere (Koch et al., 2018), we considered the novelty effect discontinuing having a notable impact on the data within the 11th week into the field study (i.e. the mean of touch down events fell below the all-year mean).

4.2.2. Phase 2

Data collection Findings at the end of *Phase 1* progressively indicated the saturation of touch interaction data. For example, as shown in Figure 6, markedly strong usage discontinued to appear in multiple instances (i.e. segments with a dark blue color and the only two-digit percentage values). In a similar vein to Glaser and Strauss (1967), who describe the process of being pointed to further data sources by the emergent theory, it felt necessary to collect additional data to enhance theoretical sensitivity. It was aimed at enriching analyses with data stemming from observations, interviews, and surveys. The motivation was fivefold:

(1) Contrary to touch interaction data, observations provide insights regarding passive usage by investigating people in their natural habitat. They place the researcher in the thick of the action (Corbin & Strauss, 2015) hence allowing to study more complex behavior (Mäkelä et al., 2018).
(2) Contrary to potential drawbacks of interviews, observations reveal what people are really doing and not what is claimed to be done (Corbin & Strauss, 2015), while being considered less disruptive (e.g. in terms of ethics) compared to, for instance, automated approaches such as video records (Parker et al., 2018).
(3) In GT, interviews and observations only provide meaning when combined (Glaser, 1998). Interview data was intended to be used as a means of contextualizing personal interpretations from observations.
(4) To truly study user experience, one must ask users for their opinions (Keskinen et al., 2013). Hence, a questionnaire was intended to be used, while they are most effective when used in conjunction with other methods (Grix, 2010).
(5) The findings of *Phase 1* were rather descriptive in nature contrary to being explanatory for what was going on. It was aimed at a more comprehensive understanding by triangulating different data sources (Grix, 2010).

Data sampling in *Phase 2* aimed at discovering the core category. In the end, data from observations of a total of 1 week (i.e. 5 working days) was gathered and enriched by data stemming from a group interview as well as a subsequent online survey. This circumstance built on the following considerations:

- It felt that a somewhat representative sample of data could be gathered during a week of observations that is sufficiently complementary to the primary data source. While we admit that further observations could have potentially unveiled more insights, we did not expect that observations such as the fact that passers-by were looking at the Ambient Surfaces almost every time would have substantially changed. The non-participant observations were conducted in Week 10 of 2015. They were not announced in advance other than in communication with persons who assisted throughout the organization process. An entire office was available and provided a direct view towards the first Ambient Surface.
- Similarly, we believed that data stemming from one group interview and a subsequent online survey would suffice to enrich the theory-generating process as these sources were also understood as complementary to the primary data source. The company internally organized for a group of representatives to participate in the semi-structured interview. It was conducted on Friday of the same week. The circa 54-minute interview was subsequently transcribed. Six people, including two Scrum Masters, three software developers, and one head of department joined the interview. The online survey followed roughly nine months later. People were able to participate between the 7th of December 2015 and the 7th of January 2016. Overall, 35 out of 76 invited employees successfully completed the self-administered and internet-mediated questionnaire (i.e. circa 46%). We recognize that conducting, for instance, additional interviews could have possibly distilled additional results. However, as a rather diverse participant group was represented in both the group interview and in the online survey, we concluded that the collected data already contained rich insights.
- The core category of *Spontaneous utilization* conceptually emerged somewhat automatically throughout the group interview. In combination with the sub-core categories of *Information visibility* and *Passing-by*, the categories seemed to account for the most parsimonious but, simultaneously, the greatest variation regarding the substantive behavior (Glaser, 1998). While sorting and writing up the emergent theory, it became apparent that data was starting to get conceptually and theoretically saturated within the limits of the available data.
- Restricted time resources, however, played a crucial delimiting part throughout this process. Generally, tasks such as conducting data collection in the field, maintaining the software application, and, in parallel, preparing the deployment of the second screen were all time-consuming activities during that time period.

Data analysis *Phase 2* aimed at expanding conceptualizations by leveraging data triangulation. The directing hypotheses and some of the research questions from *Phase 1* were subjected to an examination. For instance, we hypothesized that events such as arriving at work, leaving work, or lunch breaks were reflected to some degree in patterns we observed in the touch interaction data (see Figure 6). Consequently, we were wondering about the reasons for the patterns distilled and, for example, how passive usage would look like. As we faced limited time and project resources, we followed the advice to selectively code as soon as possible (Glaser, 1978). We issued this step in *Phase 2* to pinpoint the core category. One crucial question was denoted in the context of *Phase 2*: How is the emergent theory conceptually and theoretically enriched by

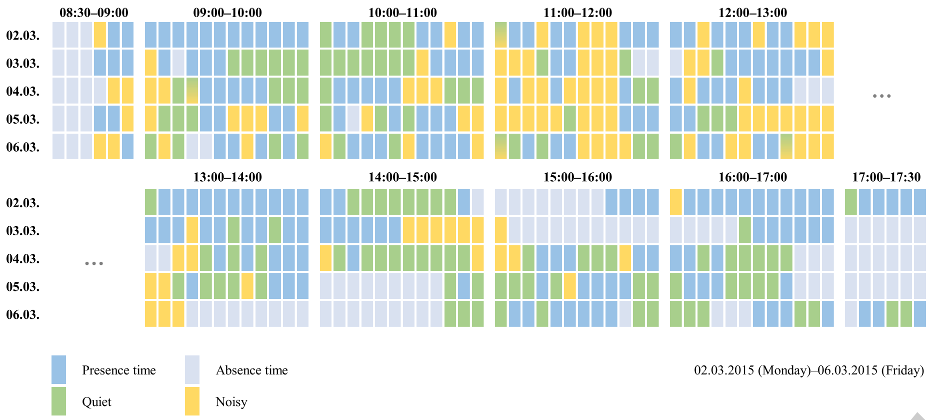

Figure 7. A memo including times the author was present (dark blue), times he was absent (light blue), as well as segments indicating noisier (yellow) and quieter (green) parts of a day.

triangulating it with observational, interview, and survey data?

Throughout *Phase 2*, open coding and selective coding practices were largely utilized. Analyses focused on identifying indicators in the data to create categories and properties. Indicators were found in phrases, sentences, or whole paragraphs which is referred to as *key point coding* (Allan, 2003). Observation field notes, the group interview transcript, and the survey results were the basis for this task. Field notes were digitally documented and analyzed using word-processing software. They contained a tabular structure that was organized in hourly segments to relate incidents to specific time windows. Events such as people passing by were documented. For analyses relating to the group interview and the online survey, *MAXQDA*[3] was used to create codes. Overall, the following comparisons were conducted:

- Compare observed noise levels (see Figure 7). It felt relevant to learn more about times of day in which people were moving more throughout the building, were talking more to each other, and—in contrast—when people were apparently concentrating more on their work.
- Compare observed times of day with respect to the total number of passers-by to elicit explanations for the varying latent patterns found in *Phase 1*.
- Compare observed instances of active and passive usage to find out more about how the first system was being utilized.
- Cross-compare findings from *Phase 1* with findings from observations.
- Compare interviewees' feedback from the group interview.
- Cross-compare findings from *Phase 1*, observations, and the group interview.
- Cross-compare findings from *Phase 1*, observations, the group interview, and the online survey.

4.2.3. Phase 3

Data collection *Phase 3* incorporated long-term touch interaction data (i.e. 2014–2017), also stemming from the Ambient Surfaces' custom touch logging mechanism. However, contrary to *Phase 1*, *Phase 3* primarily leveraged the second logging file that stored the view events. This phase embodied a response to the issue that analyses continually progress as the researcher becomes immersed in data—at some point, the

[3] `https://www.maxqda.com/`

Times of day	2014 (Weeks 18–52)	2015			2016		2017		Total	Total (in %)
		< Week 33	System 1 (> Week 33)	System 2	System 1	System 2	System 1	System 2		
7:00	9	14	20	9	0	12	0	2	66	0.18%
7:30	70	57	37	15	0	1	9	62	251	0.70%
8:00	43	106	53	22	57	15	94	121	511	1.42%
8:30	173	213	54	151	85	75	202	125	1,078	3.00%
9:00	254	348	51	96	188	123	216	252	1,528	4.25%
9:30	163	138	100	24	235	66	175	158	1,059	2.95%
10:00	306	236	92	128	256	119	180	299	1,616	4.50%
10:30	274	246	63	47	316	92	240	174	1,452	4.04%
11:00	268	483	85	120	150	81	252	268	1,707	4.75%
11:30	225	505	200	228	564	379	253	204	2,558	7.12%
12:00	297	**759**	175	**250**	**583**	354	297	292	**3,007**	**8.37%**
12:30	**653**	339	**240**	120	395	246	**306**	**406**	2,705	7.53%
13:00	267	432	192	124	281	**384**	140	231	2,051	5.71%
13:30	416	268	67	12	226	170	56	274	1,489	4.14%
14:00	213	422	62	31	387	51	255	175	1,596	4.44%
14:30	342	261	36	96	434	230	154	237	1,790	4.98%
15:00	352	424	119	39	269	166	197	187	1,753	4.88%
15:30	185	323	130	171	191	124	175	121	1,420	3.95%
16:00	285	301	33	24	302	88	114	207	1,354	3.77%
16:30	406	238	118	27	155	145	159	283	1,531	4.26%
17:00	275	195	119	41	128	61	233	264	1,316	3.66%
17:30	176	286	73	215	100	141	199	138	1,328	3.70%
18:00	113	107	57	16	58	61	59	173	644	1.79%
18:30	48	44	14	43	156	89	103	170	667	1.86%
19:00	51	51	77	48	100	73	84	83	567	1.58%
19:30	18	7	54	184	98	219	52	107	739	2.06%
20:00	4	3	38	17	23	2	0	57	144	0.40%
Total	5,886	6,806	2,359	2,298	5,737	3,567	4,204	5,070	35,927	100.00%

Figure 8. A memo showing all touch down events collected in the years 2014 to 2017 between 7:00 and 20:00 (in calendar week 33 of 2015 the second system was deployed). Times of day with the strongest usage are highlighted with a bold font and the darkest blue tone (brighter colors indicate lesser to no usage). In 2017, both systems were relocated to a new building.

researcher must stop analyzing data (Corbin & Strauss, 2015). *Phase 3* embodied a compromise of weighing up the pros (e.g. potentially gaining new insights) and cons (e.g. restricted time resources) of conducting further labor-intense data collection activities. As touch events and view events from a total of four years were readily available, we decided to use this data set to enrich the emergent theory in *Phase 3*.

Data analysis Whereas *Phase 1* largely elaborated on when utilizations occurred, *Phase 3* enriched investigations with respect to what content people were actually using. While we conducted some preliminary analyses regarding, for instance, how information views were used, in *Phase 3* we took on the task to comprehensively investigate content utilization. The aim of *Phase 3* was to further saturate the conceptual and theoretical underpinnings of the emergent theory. It was targeted at achieving theoretical completeness within the limits of the available data (i.e. all data considered from the three research phases seemed to conceptually fit). Similar to *Phase 2*, there was one analogously important question targeted in this stage of the research: How is the emergent theory conceptually and theoretically enriched by triangulating it with quantitative long-term touch interaction data? Data was retrospectively scrutinized by using spreadsheet software (see Figure 8 and Figure 9). Ultimately, the findings were cross-compared with the results of the two previous research phases.

4.2.4. Phase 4

Data collection It was decided to conclude this research with a respondent validation, which is an important corrective measure to the overall research (Torrance, 2012). It was used at the data checking stage—that is, attendees were asked whether findings were "a fair and reasonable reflection of the situation as they understand it" (Torrance, 2012, p. 114). The respondent validation was conducted on the 1st of March 2019 and organized as a group session. The nine attending staff members were: four Scrum Masters, one head of department, one computer science student, and three software developers. Three authors attended the meeting. A presentation was

Information View	2014 (Weeks 18–52)		2015 < Week 33		2015 System 1 (> Week 33)		2015 System 2 (> Week 33)		2016 System 1		2016 System 2		2017 System 1		2017 System 2	
Bug Survey View	–	–	–	–	–	–	–	–	–	–	–	–	12.00%	–	–	–
GoCD View	–	–	–	–	–	–	–	–	–	–	–	–	–	–	27.72%	↕•
Jira Views	17.01%	↕	4.54%	↕	–	–	–	–	–	–	–	–	–	–	–	–
Confluence View	15.44%	↔↕	20.49%	↔↕	87.82%	↔↕	–	–	35.43%	↔↕	–	–	30.95%	↕	–	–
Team Charts View	15.84%	↕•	12.14%	↕•	0.31%	↕•	–	–	33.64%	↕•	–	–	43.56%	↕	–	–
Test Suites View	–	–	3.29%	↔↕•	–	–	13.95%	↕•	–	–	21.87%	↔↕•	–	–	10.21%	↕•
Jenkins View(s)	8.26%	↕•	7.01%	↕•	–	–	27.82%	↕•	–	–	52.70%	↕•	–	–	12.94%	↕•
Main Window	43.39%	↔	52.50%	↔	11.87%	•	58.23%	↔•	30.42%	↔•	20.17%	•	13.46%	↔	47.83%	↔
Total	99.94%		99.97%		100.00%		100.00%		99.49%		94.74%		99.97%		98.70%	
Number of Events	164,850		186,512		41,599		14,660		125,968		29,214		119,871		70,617	

↔ = Horizontal scrolling required; ↕ = Vertical scrolling required; • = No scrolling required

Figure 9. A memo illustrating the amount of view events per information view for both Ambient Surfaces. Depending on changes in the visualizations and in the content displayed, no, horizontal, or vertical scrolling gestures were required. Some total percentage figures do not add up to 100% because some events belonged to other user interface elements (e.g. a restart button).

initially held foremost elaborating on the pictorial models of *Phase 3* and participants were subsequently asked to provide feedback in the context of an unstructured group interview. This second part was audio-recorded (roughly 43 minutes).

Data analysis Comparisons aimed at evaluating how employees contradicted or concurred with illustrations of the generated substantive theory. Consequently, the following question was targeted in *Phase 4*: How do attendees contradict or concur with the pictorial depictions of the proposed theory in *Phase 3*? The audio record of the respondent validation was transcribed and then analyzed in MAXQDA. However, coding was done with a selective lens as it was primarily concentrated on how newly gathered indicators revised the existing theory of *Phase 3*. In cross-comparing all research phases, the final round of sorting and writing up took place. This resulted in a concluding pictorial depiction (see Section 5). No further sampling of data was conducted.

5. Results

GT researchers are encouraged to freely design what they intend to present and to write about a theory in a substantive manner (Glaser, 1998). Accordingly, this section introduces the final pictorial models that evolved throughout *Phase 1* to *Phase 4*.

5.1. Overview

In *Phase 1*, findings were rather descriptive in nature (e.g. time windows with peaks in interactions), while we successively unveiled the conceptual and theoretical foundations of the emergent theory in *Phase 2*. Ultimately, it was realized that utilization was largely linked to spontaneous incidents and that the theory consists of three main pillars to describe the substantive behavior of staff members. While people were reporting their feedback, they repeatedly highlighted aspects concerning the visibility of information (e.g. the relevance of certain information), the process of walking past the Ambient Surfaces (e.g. to fetch a beverage), and finally, their ad-hoc utilization behavior (e.g. when arriving at work). Consequently, the categories *Information visibility*, *Passing-by*, and *Spontaneous utilization* were created (see Figure 10). In *Phase 3* it was observed, for instance, that the desired information differed throughout a day, while the overall set of information views had an emphasis on progress tracking. Throughout

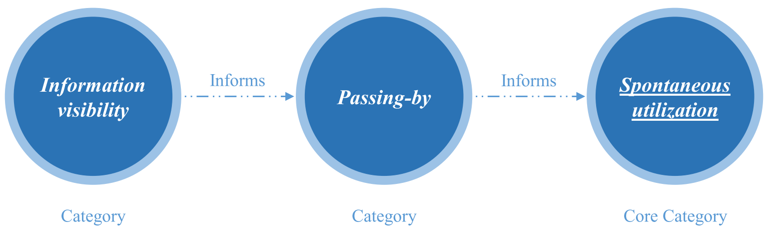

Figure 10. The three main categories of the proposed theory.

Phase 4, we experienced people anecdotally reporting from daily routines where the Ambient Surfaces were involved (e.g. while arriving at work).

In the following figures, both the category of *Information visibility* and *Passing-by* precede the core category of *Spontaneous utilization* as they were found to inform this center piece of the proposed theory. The logical order is from left (the category of *Information visibility*) to right (the core category of *Spontaneous utilization*). For the sake of presentation, different visual means were utilized. While the three central categories are predominately represented in the figures, categories and their properties are shown as smaller rectangle shapes. There are solid lines and three types of arrows included in the illustrations. All arrows are equipped with corresponding labels. In the case of solid arrows, these labels are a product of self-generated aggregating codes, but they do not stem directly from the data—they are an effort to encapsulate the "not obviously stated" (Glaser, 1978, p. 56). While solid lines connect categories with properties, the three types of arrows are leveraged to: firstly, point to aggregating theoretical codes (solid arrows); secondly, highlight the varying theoretical codes that emerged throughout analyses (dashed arrows); and thirdly, indicate the varying paths through the theory (both dotted and dashed arrows). The idea behind highlighting theses paths is to illustrate under what circumstances utilizations occurred (four cases), what prevented staff members from engaging with the Ambient Surfaces (six cases), and what were the observed outcomes of utilization incidents (three cases).

5.2. The category of Information visibility

This category is built upon feedback surrounding the visibility of information (see Figure 11). It explains why—or why not—people utilized the Ambient Surfaces. The category is affected by different reported *Impacting factors*. While the continuous *Evolution* of the system helped in upholding a current state of information displayed (e.g. by adding a new information view), *User interface issues* such as the amount of simultaneously displayed information were perceived as challenging. For example, regarding the *Jira* view one employee said that the teams had to "try to compete for some of the entries on top" to get recognition. A critical issue was also the availability of just one Ambient Surface prior to August 2015 (i.e. the *Number of displays*). One employee stated that "the screen could be at least five times as large" to mitigate the issue of not seeing information due to the limited available space on one screen. Initially, the user interface was 2.5 times the size of a 1,080p resolution in width and people had to notably navigate through the content to unveil desired information. Principally, "you can't have enough displays" as on employee vividly said. Furthermore, the *Format* of the Ambient Surfaces itself affected the visibility of information since "information is presented in a different way outside the normal access points." Both *Locations* were appreciated in the feedback. For instance, one employee characterized the first setup as a "meeting point", where people reportedly gathered after fetching a coffee from the kitchen. The two locations were also found to affect patterns of passers-by.

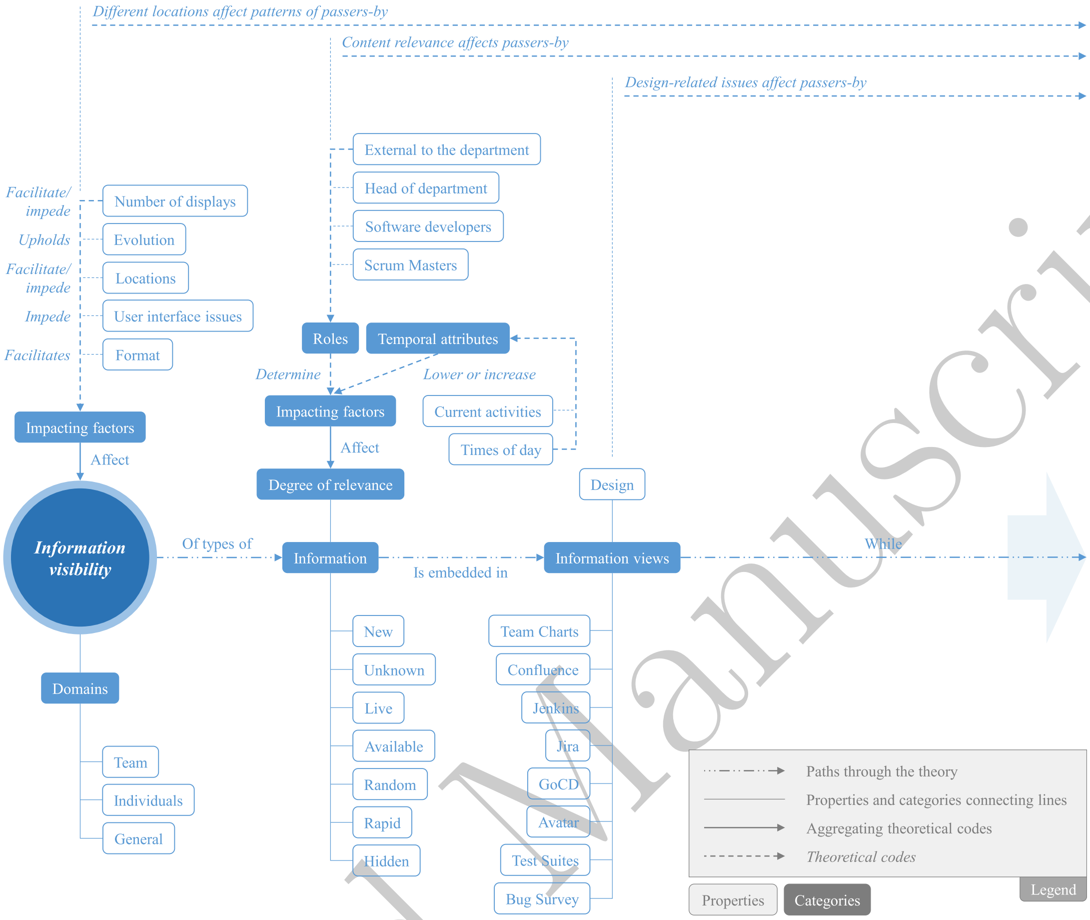

Figure 11. The pictorial model of the category of *Information visibility*.

People referred to the visibility of information in different contexts, denoted as *Domains* (i.e. *Team*, *Individual*, and *General*). While some talked about team-related issues (e.g. *Team Charts* view), others mentioned individual (e.g. *Jira* view) and more general (e.g. visibility of available information) aspects. The Ambient Surfaces created visibility of different types of *Information*. In sum, the feedback resulted in seven different adjectives that characterize information further (e.g. *New*, *Live*, or *Random*). In addition, it was found that information has a *Degree of relevance*, which affected people while they were passing by, since it impacted the decision-making process whether to engage further. It could be observed that both the *Roles* of individuals and *Temporal attributes* influence the *Degree of relevance*. While, for instance, it was reported that views such as the *Jenkins* view are "more interesting for the teams", Scrum Masters and one head of department noted a preference towards visualizations such as the *Team Charts* view (i.e. more overview-orientated). Furthermore, to personnel external who are "not directly involved in the development process", the Ambient Surfaces reportedly provided assistance. Regarding *Temporal attributes*, it could be revealed that the varying views showed strong utilizations during different *Times of day*. For example, while the *Confluence* view showed peaks in interaction at 13:30 and 15:30, the *Team Charts* view indicated similar peaks at 11:30 and 17:00. Additionally, *Cur-*

rent activities (e.g. a Sprint end) in the ASD department can affect the relevance of certain information (e.g. burn down charts). These activities should be considered in displaying information as systems would be then "much more likely to be used."

Finally, the displayed *Information* was embedded in *Information views*. In total, eight different views were created. Each of them had an individual *Design* that affected passers-by as well. For instance, views that required interactions to unveil information (e.g. the *Confluence* view), showed measurable differences during analyses compared to views that at times did not require any interaction (e.g. the *Team Charts* view). This issue was most notable in 2015 for *System 1* (see Figure 9). While the *Confluence* view displayed long articles that necessitated scrolling gestures, the *Team Charts* view illustrated plain chart diagrams that often required no such interactions.

5.3. The category of Passing-by

This category summarizes feedback regarding the process of passing by the Ambient Surfaces (see Figure 12). It elaborates further on the question of why—or why not—people engaged with the Ambient Surfaces by presenting the procedures included that led to or prevented the systems from utilization. The category represents a link between both other categories and was repeatedly indicated in the feedback. Principally, content was presented to *Individuals* and *Groups of individuals* while they were passing by the systems. People potentially paid no further attention towards the Ambient Surfaces in five instances: firstly, when they were not facing the displays directly while walking past them; secondly, when the systems were already in use; thirdly, when they were in a rush or in a hurry; fourthly, when they were accustomed to acquire information elsewhere; and lastly, when they typically ignored the systems due to not expecting any interesting information (i.e. display blindness).

The process of passing-by was affected by five *Impacting factors*. Firstly, the total number of passers-by, the ratio of individuals and groups of individuals, as well as the audience type all depended on what is referred to as *Times of day*. Participants recognized, for instance, that during the canteen opening hours (i.e. 11:30–14:00) in the second setup, the audience was "kind of mixed", while beyond this time frame the audience consisted of mostly ASD department staff members. Furthermore, it was observed that during lunch breaks, more groups of people gathered around the Ambient Surfaces—the time period with the strongest peaks in interactions (see Figure 8). Secondly, *Display blindness* lowered the perception of the systems hence affected the tendency to pay attention towards them while passing by. While some people expectedly "ignore it [the Ambient Surfaces] completely", display blindness may have also incrementally developed due to reasons found in repeatedly not seeing desired information. Thirdly, how people conducted their *Information work* affected the perception as some people preferably acquired information elsewhere. Fourthly, the *Facing position* towards the systems was found to affect utilizations. Were people directly facing the systems, it was more likely that the systems were used. Finally, *System vacancy* potentially influenced the willingness to consider stopping in front of the systems, because if "there are already three people in front of it [the Ambient Surfaces], then I pass by, because I don't have the chance to interact at all."

When people were looking at the screens (i.e. they became somewhat aware of the systems and the scene in front of them), they were seemingly involved in a procedure that is denoted as *Pre-engagement process*. During this process, a decision to engage further was seemingly concluded in a split second. The process accounts for the ap-

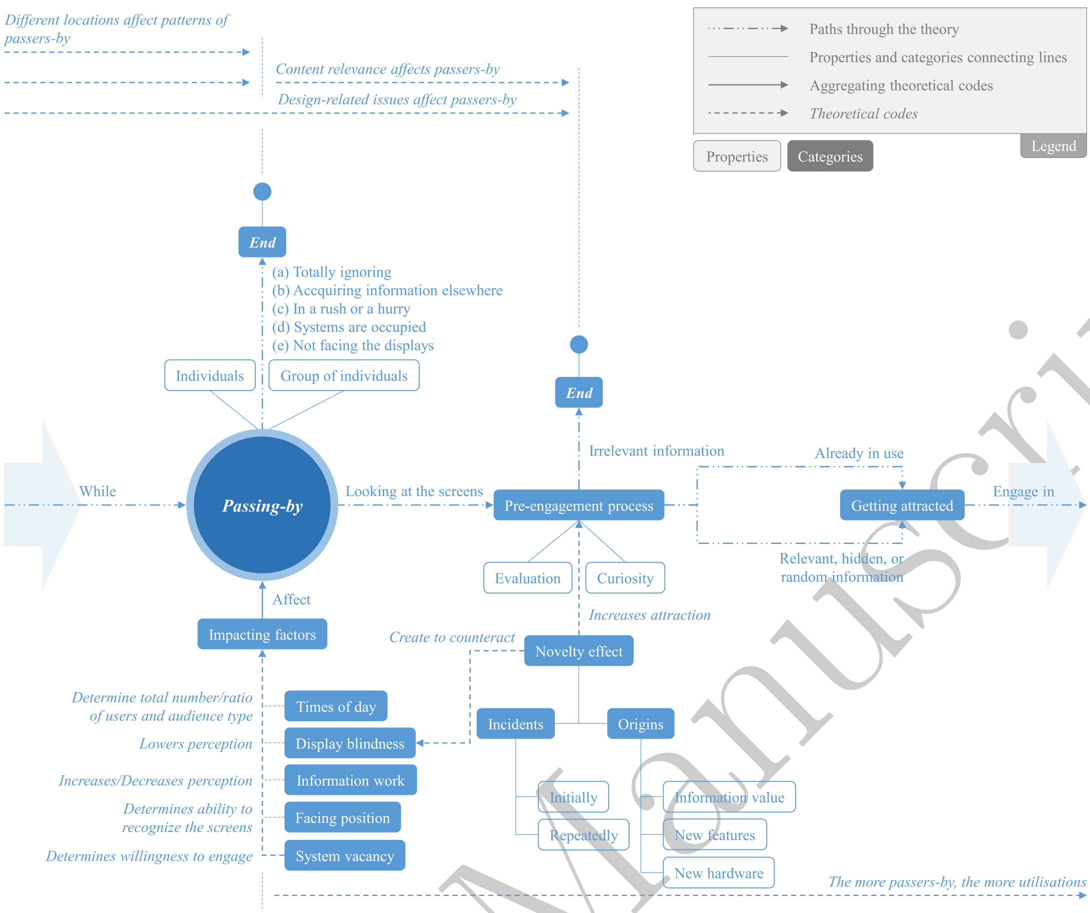

Figure 12. The pictorial model of the category of *Passing-by*.

parent event that people were driven by, on the one hand, an *Evaluation* of a piece of information's relevance (i.e. relevant or irrelevant information) or visibility (i.e. hidden information) and, on the other hand, a *Curiosity* stemming from seeing random information or the fact that someone was already standing in front of the systems. This latter phenomenon is known as the *honeypot effect* (Brignull & Rogers, 2003). A participant anecdotally illustrated the issue of curiosity in saying: "Oh look, he [a colleague] just put something there [the *Confluence* view] ..., what is it?" In terms of an *Evaluation*, information relevance was crucial. When information was displayed "that doesn't interest me in my daily work right now, then of course I won't look any further." In these and similar cases, people simply passed by the Ambient Surfaces. In summary, people were *Getting attracted* in cases of relevant, hidden, or random information as well as when the systems were already in use. A comment from one employee underlines this summary vividly: "I would miss it [the Ambient Surfaces] too, because this getting random information that is somehow relevant ... and related to the work—that's actually quite nice." The *Pre-engagement process* was furthermore affected by the *Novelty effect*, which potentially increased the attraction towards the systems. Two *Incidents* were revealed in which the effect occurred: *Initially* (i.e. at the study's commencement) and *Repeatedly* (e.g. when new updates were deployed). Analyses also led to the identification of this effect's *Origins*: firstly, *New features* (i.e. after deploying new features); secondly, *New hardware* (i.e. after deploying the second screen in August 2015), and finally, *Information value* (e.g. people supposedly tended to investigate content more frequently when "bad news" such as broken builds was presented). To counteract display blindness (or the "wear and tear effect" as someone commented), participants also introduced the idea of repeatedly leveraging the novelty effect in terms of "that you always bring up something new".

As indicated in the previous section, an information's *Degree of relevance* was found to affect passers-by. This is accounted for to the following extent in the theory. Information relevance can respectively affect both the *Pre-engagement process* (i.e. information was considered relevant or irrelevant) and, prior to that, the process of *Passing-by* itself as regard to the magnitude of *Display blindness* (i.e. people partially or totally ignored the systems). Finally, cross-comparisons of *Phase 1* and *Phase 2* suggested a correlation between the total number of passers-by and utilizations. For instance, increases and decreases in the number of observed passers-by (*Phase 2*) were reflected in patterns of touch interaction data (*Phase 1*).

5.4. The core category of Spontaneous utilization

This category casts light on how study participants described their actual utilization behavior (see Figure 13). It expands on mechanisms that are related to utilization and highlights the very nature of such incidents. The feedback repeatedly distilled the spontaneous nature of utilization, strongly linked to the process of *Passing-by*. One statement vividly reflects a common theme in the data as to, evidently, people utilized the Ambient Surfaces in "spontaneous occasions when I walk past it." Reported incidents of ad-hoc situations where, for example, fetching a beverage from the kitchen, arriving at work, or entering the building. The two locations facilitated impromptu situations, because "every colleague passes by there several times a day." Rarely, indicators for deliberate use were found in the data. Employees noted that, for instance, "developers do not use this [the Ambient Surfaces] consciously" or "I rarely go to the computer [the Ambient Surfaces] purposely and look at something." Another participant explained that he or she "wouldn't even know how to do something somehow planned" with the systems. Supposedly, consuming information at the desktop computer is "much easier", "more accessible", and "more pleasant". It is the summary of this feedback that led up to the core category's name.

As said before, when people were attracted by, for instance, relevant or random information, they engaged in these spontaneous incidents of utilization. Fundamentally, utilization was reported to be affected by *Individual preferences* (e.g. active or passive usage). Supposedly, the systems were utilized "very individually" beyond a particular role someone holds (e.g. Scrum Master). When considering the ratio of individual users in contrast to group usage, a tendency of use cases towards single users was observed. One staff member was comparing the Ambient Surfaces to a coffee machine, meaning that it attracts people "and when people stand in front of it, it attracts even more" (i.e. the honeypot effect). Another employee compared it to a traffic light—when one person crosses the street through a red light, all others follow.

The three most popular mentioned reasons to utilize the Ambient Surfaces were gaining an initial overview, catching up on project information, and staying informed of colleagues' activities. Principally, the *Motivation* to utilize the systems was found to be threefold. Firstly, people were foremost referring to aspects relating to *Awareness* such as team awareness. As one employee said: "Transparency through the burn downs [the *Team Charts* view] -> you can quickly ask your colleagues in the other teams what

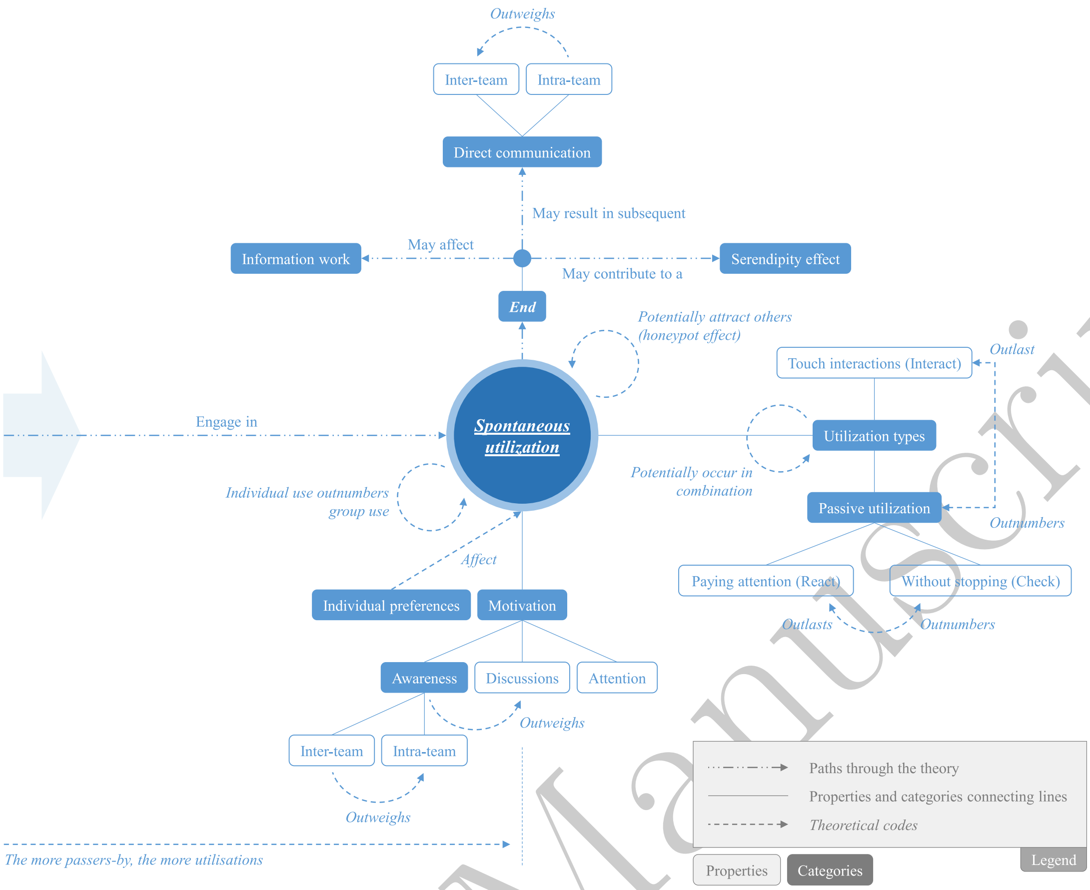

Figure 13. The pictorial model of the core category of *Spontaneous utilization*.

is going very well or very badly." It could be found that the Ambient Surfaces promoted a better *Inter-team* awareness than *Intra-team* awareness. A comment from a staff member describes that issue vividly: "The Ambient Surface system is not bringing any new information from my team. ... Through internal dailies and spatial proximity, you are better informed." Secondly, people leveraged the systems to also be involved more frequently in *Discussions* with colleagues. However, the relevance of awareness-related issues was found to outweigh the importance of being more often involved in discussions. Thirdly, the Ambient Surfaces were also used to draw *Attention* of others to specific content as a participant described. Reportedly, for him or her new articles in the *Confluence* view were "often a reason for me to stop" to indicate to others that "here is a new blog post" (i.e. to purposely cause the honeypot effect). In his or her view, the responsibility to make the systems attractive lied in the staff's own hands.

Three types of utilization were revealed (denoted as *Interact*, *React*, and *Check*). While *Interact* refers to incidents where people were physically interacting with the Ambient Surfaces, *React* points to situations where people became attracted by content in such a way that they focused their attention on the screens (e.g. they stopped in front of the systems). *Check* denotes incidents where people passed by and looked briefly at the Ambient Surfaces, but they continued their way without stopping. While *Touch interactions* (i.e. *Interact*) typically outlasted *Passive utilization* (i.e. *React* and *Check*), the latter outnumbered the former. While *Check* utilizations potentially accounted for most incidents, they were simultaneously the briefest in nature. Yet, *React* utilizations outlasted incidents of *Check*. All three types may also occur in combination. For example, it was repeatedly observed that people were walking past the Ambient Surfaces, but suddenly stopped for a second (i.e. a *Check* utilization transcended into a *React* one) and, sometimes, they then physically interacted with the screens (i.e. a transition to *Interact*).

Finally, utilization implications could be identified. Respondents reported that they proactively confronted colleagues in subsequent *Direct communication* due to displayed information and, simultaneously, were confronted because of visualized content. For example, a staff member reported that when he or she saw an issue on the screens "where you somehow knew something about ... or just wanted to have further information, you just looked for the colleagues who were really working on it." Another employee experienced situations where he or she was proactively confronted by colleagues with the question: "Why did the burn down [the *Team Charts* view] went up?" Interestingly, communication and awareness findings were contradictory in the sense that predominantly *Intra-team* communication over *Inter-team* communication was encouraged. Additionally, there were changes observable in people's way of conducting *Information work*. For instance, one respondent said that the systems reminded him or her to subsequently investigate information further. Some employees also started to exclusively consume selected information on the Ambient Surfaces. As a participant noted: "I no longer obtain some information from the workstation computer, as I have already received it through the Ambient Surface." Another participant stated that he or she was in 90% of the cases obtaining team charts information solely on the Ambient Surfaces. Finally, evidence was pinpointing the *Serendipity effect* (Ott & Koch, 2012) as to people also discovered relevant information by accident. Reportedly, "you can even see articles that you would not otherwise look at."

6. Discussion

We fundamentally followed the stance from Bryant (2009), where pragmatists are described as having a focus on continually generating new ideas about the world. For pragmatists, this process is never really completed. Similarly, classic GT presents probability statements that are readily modifiable as new data emerges (Glaser, 1992).

6.1. Originality

According to Glaser (1998), every generated theory in classic GT methodology has strong contribution strength. We reflect on the originality of our work considering three central topics: tools in the software vitalization domain, behavioral models and social dynamics, as well as the ecological validity of the data.

6.1.1. Tools in the software visualization domain

To the best of our knowledge, the originality of the proposed theory builds on two main pillars. Firstly, a central characteristic of the Ambient Surfaces solution is providing information to the entire ASD team. Contrary to many related tools, it does also not focus on a single data source as this could be incomplete or insufficient (Ye et al., 2018). Similarly, our solution is different from many other tools as it does not emphasize on a single-user perspective which breaks with the typical metaphor of developing ambient awareness tools for the scope of a desktop computer (Sharma et al., 2019). Secondly, the theory elaborates on novel insights in the under-researched area of long-term ambient display deployments in professional ASD contexts (Schwarzer et al., 2016). There is a crucial need for empirical research regarding software visualization (Bedu et al., 2019), while a focus on industrial contexts (Mattila et al., 2016) and on authentic target audiences (Merino et al., 2018) is suggested.

6.1.2. Of behavioral models and social dynamics

The theory may assist in developing still lacking behavioral models of people facing such technology by distilling the social dynamics (Ardito et al., 2015). Understanding how people approach display technology, what issues lead to non-use or avoidance, and how people engage in collaboration, are crucial factors to investigate in ambient display research (Ardito et al., 2015; Williamson & Williamson, 2014). The core category of *Spontaneous utilization* reflects a common phenomenon that is occurring, when informal awareness, as a prerequisite for spontaneous interaction (Gutwin, Greenberg, & Roseman, 1996), is created: coordination and collaboration are facilitated as well as ad-hoc set ups for social exchanges are initiated (Gross, Stary, & Totter, 2005). Related studies designed whole systems considering this notion (e.g. Rittenbruch, 2016) and, similarly, found that spontaneous social debates between study attendees were encouraged (e.g. Rubegni, Memarovic, & Langheinrich, 2011). Like Markopoulos (2009), who reports that awareness can provide the trigger for communication, we observed conversations in front of the screens and, simultaneously, staff members reported that they subsequently confronted colleagues due to displayed information. Another facet of the observed social dynamics stems from both the composition of the group of potential users and the individual commitment to increase the value of the Ambient Surfaces solution. Contrary to many other ambient display studies (Mäkelä, Sharma, Hakulinen, Heimonen, & Turunen, 2017), we considered staff members not as simple passers-by, but took their individual roles into account throughout analyses. Compared to Parker et al. (2020), we also found that the background of users changed throughout the entire day and in the two locations. We add to this comparison that information views provide different levels of value, dependent on the times of day. For example, the *Confluence* view was used the most at around 13:30 and 15:30. We concur with Parker et al. (2020), who report from a community's own opportunity to impact the value of a system, as we could reveal that some employees proactively used the Ambient Surfaces to make others aware of new information.

Since there is a lack of existing general theories regarding ambient displays (Alt et al., 2012), the implications of our work can be considered substantial as they point to conceptual and theoretical building blocks that may assist in developing behavioral models in the future. The theory may help with creating a still missing theoretical understanding regarding the psychology of interactivity (Rubegni et al., 2011). Our work brings forth insights concerning, for instance, why staff members refused to utilize the Ambient Surfaces (e.g. consuming information elsewhere) or why they engaged with the systems (e.g. seeing relevant information). It demonstrates how common phenomena such as the novelty effect and the honeypot effect occurred. Overall, our work unveils how the systems were accepted in real life (Colley et al., 2018).

Our findings stand in some contrast to the work from Ghare et al. (2018), who's results indicate that the strongest motivation to approach their display solution lied in the honeypot effect, while our research yields awareness-related reasons as well (e.g. gaining an initial overview). The honeypot effect, as the most scrutinized impact factor on audience behavior (Mäkelä et al., 2020), seemed to not play such a prominent role in our research. Further differences to related studies became apparent, when considering the so-called *audience funnel* framework (Michelis & Müller, 2011). While, for example, Ghare et al. (2018) differentiate five behavioral codes (i.e. *none*, *glanced*, *stopped*, *explored*, and *touched*) that show how people interact with a display, we summarized this behavior in three such codes (i.e. *Check*, *React*, and *Interact*). We had come to observe that people glanced at the screens almost every time they passed by (i.e. *Check*) hence we barely saw evidence that staff members paid no attention (i.e. code *none*) towards the systems. These very brief moments constituted most cases, while it is typical to see more passive than active users (Mäkelä et al., 2018).

6.1.3. Ecological validity of the data

Contrary to related studies (Ghare et al., 2018), the Ambient Surfaces neither went both out of use nor unnoticed at some point in time. Typically, the number of users drops over time (Mäkelä et al., 2018), while we observed notable amounts of touch interaction data over the years. Following Hazlewood et al. (2011), we think that our solution became somewhat truly ambient at Werum and showed features of calm technology. Taking into account that staff members kept utilizing both systems throughout the entire study—hence adopted them to a lesser or greater extent—and, simultaneously, that tool adoption is considered the strongest evidence of usability (Alves, Niu, Alves, & Valença, 2010), we conclude that the original contribution is built on data showing high ecological validity (Alt et al., 2012). For example, we carefully considered the novelty effect throughout analyses—a threat to this validity that we recently investigated (Koch et al., 2018). Specifically, we did not consider touch interaction data from the first weeks of the deployment in 2014 and we also repeatedly removed data after the Ambient Surfaces had been updated (e.g. a new information view). In doing so, we avoided, for instance, data in *Phase 1* skewing conclusions regarding usage at the varying times of day. Another example is the familiarity of employees with the content displayed. Opposed to leveraging attention-grabbing visualizations, our solution was developed to show authentic, well-known information from the department.

6.2. Scientific usefulness

Scientific usefulness largely derives from implications that are exemplified by the continuous evolution of the Ambient Surfaces and that are manifested in the methodological elaborations in this study.

6.2.1. The interplay of environment and technology

Similar to Williamson and Williamson (2014), who report from a reciprocal relationship between the environment and technology, we also experienced this phenomenon in our research. Both Ambient Surfaces regularly underwent minor (e.g. font recoloring) and major (e.g. removing or adding a view) modifications to account better for the staff's demand. For example, the staff specifically developed new tools to be displayed on the Ambient Surfaces (e.g. the charts used in the *Avatar* view). We observed this interdependent relationship to varying degrees throughout the entire study and see parallels to what Crabtree, Hemmings, and Rodden (2003) refer to as *situated displays* and *coordinate displays*. Situated displays describe socially constructed artifacts that people create throughout their routine interactions. Coordinate displays are a distinct class of situated displays that target at collaborative work. Both types of displays are routinely constructed and continuously shaped, while we think that our solution incorporated more the notion of coordinate displays. In essence, in long-term research we might be urged to reflect on a system's life cycle (Brown et al., 2017) and our experience underlines the necessity to do so. The evolutionary development of the Ambient Surfaces (see Figure 2), in our view, prompts researchers to rethink their methodological approaches, when it comes to scrutinizing technology in the field. For example, it challenges prevalent quantitative experimental designs with respect to the evaluation of ambient displays in the software visualization context (Mattila et al., 2016). As both Ambient Surfaces were continually revised, team organizations evolved, and locations changed, it was arguably impractical to go beyond the means of abductive inquiry (e.g. to use parametric tests with fixed variables).

6.2.2. Methodological considerations

To really grasp on how technology is used in practice, longitudinal research as well as a cross-cutting through both qualitative and quantitative data sets is required (Jurmu et al., 2016). As Shneiderman et al. (2016) summarize, conducting research in the wild requires improved interdisciplinary methods. We share this view and therefore see fundamental implications for research that seeks to provide empirical and theoretical contributions in this domain. As long-term in-situ research on ambient displays is increasing, there is a need for improved evaluation methods (Mäkelä et al., 2018). In-the-wild studies of ambient displays are difficult to control, while also being fragile with respect to intervention bias (Williamson & Williamson, 2017). Our answer to this challenging task was the selection of classic GT since this methodology is said to reconcile bias of people and methods in the process of generating theory (Glaser & Strauss, 1967). Throughout *Phase 1* to *Phase 4*, the pictorial models incrementally developed and assisted greatly in keeping track of the theoretical and conceptual building blocks of the emergent theory. We extensively leveraged memos to generate new ideas about the data and to reflect on them collaboratively. We find the methodology particularly helpful, because it emphasizes on the conceptual and theoretical development of data rather than imposing on researchers to choose between quantitative and qualitative data sources. A discussion of good and bad data becomes useless in GT methodology (Glaser, 1998). We concur, for instance, with Mäkelä et al. (2018), who conclude that observations and automated logging can be supportive to one another in multiple ways. Thus and similar to Parker et al. (2018), we decided for a mixed-methods design, while we did not plan in advance the chronology of data collection procedures other than to use touch interaction data as the primary data source. The constant comparison process pointed us towards insightful, new data sources as the research progressed.

Yet, as we recently expanded upon (Schwarzer et al., 2019), we are not of the opinion that our chosen way may fit every in-situ research the best. Sometimes, researchers may have to develop their own methodological approaches (Keskinen et al., 2013). However, this study responds to and concurs with recent developments in HCI and CSCW research that encourage a more practice-based research agenda (Jurmu et al., 2016). We adopted these developments by bringing an existing methodology capable of studying social phenomena to a field where it is, admittedly, still not clear what constitutes its foundations (Muller, 2014). Nonetheless, the application of classic GT methodology may prospectively assist in resolving some issues reported in ambient display in-the-wild research such as the transferability of findings (Mäkelä et al., 2020)—i.e. existing and new methods need to scale up beyond the investigated specific setting (Brown et al., 2017). Fundamentally, every generated theory has its own degree of substantive or formal generality (Glaser, 1998). In this light, the present study may indicate directions to cope with, one the one hand, the need to personalize ambient display solutions, and, on the other hand, to generalize findings to other contexts (Ghare et al., 2018). We understand classic GT as a promising means to contribute a more "systematic framework" (Ardito et al., 2015, p. 28) for in-the-wild evaluations that accounts for the particular aspects of ambient displays (e.g. environmental factors). Principally, Shneiderman et al. (2016) invite a fresh thinking when conducting research on sociotechnical systems. New assessments methods and tools are required as well as evaluation methodologies to consider all aspects of the user experience (Stephanidis et al., 2019). As GT methodology encourages to embark a study without knowing the problem up front (Glaser, 1998), we believe that it inherently shows great capabilities to advance on this development. In being methodologically explicit as to how we went about in our research, we hope that other researchers will find fruitful ideas to design their own long-term evaluations of ambient displays in the wild.

6.3. Practical usefulness

This research similarly shows practical usefulness. Here, we focus on the mentioned issues of knowledge sharing and inter-team coordination in large-scale and very large-scale ASD environments (Dingsøyr et al., 2018). Different fora for coordinating and sharing purposes such as the Scrum of Scrums meeting are used, yet these meetings do not necessarily lead to satisfactory coordination (Paasivaara, Lassenius, & Heikkilä, 2012). We expect agile practitioners to find the most value in two aspects from this research: the topic of social connectedness and the evolved set of information views.

6.3.1. Social connectedness

Fundamentally, work group members need to know about one another as well as about shared artifacts and group processes (Gross et al., 2005). Social connectedness, as one of the strongest motivational factors for the use of awareness systems, exemplifies the principal human need for awareness which is deeply ingrained in us as social beings (Markopoulos, 2009). We concur with Parker et al. (2018) with respect to seeing the importance of contextually relevant and physically embedded information, particularly in the proximity of people's main paths. Overall, the Ambient Surfaces' continuous utilization at Werum arguably reflects the desire of being socially connected. We consequently assume that agile practitioners will likely utilize such solutions when they are deployed in their surrounding environment providing they are continuously maintained to account for the corresponding demand. Among others, the present study shows when to expect, for instance, the most interactions (i.e. around lunch time) as well as what information had a particular relevance during the different times of day. Agile practitioners may derive suggestions from that as to, for example, how to better prioritize and leverage existing dashboards and information radiators.

6.3.2. The evolved set of information views

Considering that the visual design of software visualization tools is a recurrent issue (Bedu et al., 2019), we can report that we largely concur with Shneiderman's often times cited statement that says "[o]verview first, zoom and filter, then show details on demand" (Shneiderman, 1996, p. 336). All eight information views were analogously created. However, we generally feel that the most important finding from this set of views is that it implicates the priority of certain information. The Ambient Surfaces in their latest version contained views regarding team progress, bug fixing progress, continuous delivery progress, test metrics, and an overview of test outcomes and ownerships. To summarize, we saw a strong emphasis on progress tracking, which is also reflected by studies in this field (e.g. Paredes et al., 2014). Understanding which visualizations are most effective, is valuable for designers to create collaborative agile tools in the future (Scott-Hill et al., 2020). Especially in aspects embodied in the category of *Information visibility*, agile practitioners may find guidance for the difficult task of designing seamlessly integrated ubiquitous applications (Marquardt & Greenberg, 2015). For instance, this study envisions directions to handle architectural decisions in large-scale ASD environments, which are typically not supported by agile tools out of the box (Nord, Ozkaya, & Kruchten, 2014). Employees at Werum shared such decisions in the *Confluence* view via detailed articles.

6.4. Scope

Because conceptualization in classic GT methodology goes beyond place and time, Glaser (1998) argues that every generated theory has its own degree of generality. In classic GT, the theory is not generalized to a unit, but to the core category. A core category can be readily transferred to another substantive area to, for instance, reach substantive generality—or to finally state formal theory. Fundamentally, this procedure is one of the empowering features of classic GT. Nonetheless, there is a diversity of existing formats of ambient displays and a multitude of varying contexts of use (Börner et al., 2013). Like Rittenbruch (2016), however, who discuss transferability properties of their large display solution, we believe that our theoretical contribution can also be conceptually applied to both comparable ambient display solutions as well as similar large-scale and very large-scale ASD environments.

6.5. Limitations

While we followed Glaser (1998) in using the essential books on classic GT (i.e. Glaser, 1978, 1992; Glaser & Strauss, 1967), we—as novel GT researchers—do see the possibility that we may have misinterpreted elements of the methodology. Similar to Keskinen et al. (2013), however, we do not claim that our methodological approach is complete. Like Jurmu et al. (2016), we consider our systematization of classic GT as a first step to rigorously investigate ambient displays in the wild. It should also be considered that the application of classic GT is a highly individual endeavor hence its result is a personal perspective informed by an individual's cognitive style (Morse et al., 2009). This study is also limited in terms of its focus on one substantive area and on one specific type of ambient displays that leverages one's vision and haptic senses. A limitation additionally arises from considering touch interaction data as the primary data source hence passive utilization was not investigated to an equal extent. Restricted time and project resources were the primary delimiters in the process of data saturation and resulted in this emphasis. We acknowledge that additional in-field observations and interviews could have unveiled further insights.

Coping with bias was an important task for us. In the interviews and the online survey, we emphasized that anonymity and confidentiality of personal data would be maintained, while we encouraged both supporting and contradicting results. During the interviews, we tried to, for instance, avoid speaking competitively to maintain a neutral position as moderators. For the online survey, we initially conducted a pilot test and collaborated with a university department specializing in survey research, which also provided the utilized online survey tool. We aimed at samples including people of different roles in both the interviews and the online survey. During observations, we coped with observer bias by critically reflecting on field notes after a day in the field. Ultimately, we collaboratively discussed any empirical and theoretical findings.

6.6. Recommendations for future research

In our recommendations, we concentrate on four central topics. Firstly, we conclude that there is still a considerable gap of existing knowledge with respect to long-term evaluations of ambient displays in real-world environments such as ASD contexts. We have yet to reach a sound understanding on how these devices are utilized by agile practitioners. We invite future studies to contribute towards this understanding by, for instance, transferring the proposed theory to other areas to reach substantive generality (e.g. very large-scale ASD environments with more than 10 agile teams). Ultimately, the goal would be to state formal theory at some point and, consequently, to propose general theories on these devices. Secondly, like Mäkelä et al. (2018), we argue for improved evaluation methodologies for longitudinal in-the-wild research of ambient displays. There is still a need for methodological development (Du et al., 2017) and there are existing epistemological (Jurmu et al., 2016) and methodological challenges (Schwarzer et al., 2019). Thirdly, we draw attention to methodological development as regard to GT methodology itself. The methodology shows fruitful directions for future development since its application is still considered uneven or even schismatic in HCI and CSCW research (Muller, 2014). Simultaneously, more GT studies are required that demonstrate leveraging a mixed-methods approach (Walsh, 2015).

Finally, we call attention to recent tool developments in aiding the investigation of ambient displays in the wild (e.g. Elhart et al., 2017; Mäkelä et al., 2018). To better grasp the user behavior, these studies recommend using camera-based sensors. Studies based on this tool development would evidently also portray insightful research. We already collected first material with two Microsoft Kinect cameras in the past (see Figure 1). While we did not include this data in the process of generating theory here, we are planning to do so in the future.

7. Conclusion

Longitudinal ambient display research in authentic environments is still scant and demands methodological development. Epistemologically and methodologically, we responded to this situation by commencing our study rather openly. We did neither define any hypotheses or research questions in advance nor were guided by the strict means of deductive or inductive inquiry. Rather, we were adopting a pragmatic and abductive lens aiming at finding the best possible explanations we could find in the data. In this vein, we deployed two Ambient Surfaces in the large-scale ASD department of a German company and were able to operate them for approximately 5 years. Both systems displayed familiar information from the ASD department, while targeting Scrum Masters, Product Owners, software developers, and the management alike.

We chose to apply classic GT methodology in our research and incrementally cross-compared ecologically valid quantitative and qualitative data throughout four research phases. Successively, these phases built up the original, theoretical contribution of this work—a substantive theory that shows how professional agile practitioners utilize ambient displays long term. Over time, we learned that utilizations arose primarily from spontaneous occasions. The core category of *Spontaneous utilization* reflects this overriding pattern found in the data, while indicating subordinate categories and properties as well as theoretical relationships. Fundamentally, the included information views developed an emphasis on progress tacking, whereas the solution's strength could be ultimately seen in providing inter-team awareness to employees and encouraging intra-team communication among them. The systems further caused staff members to draw others' attention to information and they were continually incorporated in daily routines such as during lunch breaks. Evidently, the Ambient Surfaces solution became a somewhat integral part of the ASD department's environment.

Prospectively, longitudinal studies such as this one, may assist in creating a better understanding of how ambient displays are used in real life—and by agile practitioners in particular. Both more empirical knowledge as well as methodological foundations are necessary to, on the one hand, understand their actual value to users and, on the other hand, equip future evaluations with adequate scientific rigor. We hope that with its rich methodological, theoretical, and practical elaborations, other scientists will find our study helpful to embark on this challenging task in the future.

Acknowledgments

We would like to thank Werum and the ASD department for their long-standing participation in this research.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Allan, G. (2003). A critique of using grounded theory as a research method. *Electronic Journal of Business Research Methods*, *2*(1), 1–10.

Alt, F., Schneegaß, S., Schmidt, A., Müller, J., & Memarovic, N. (2012). How to evaluate public displays. In *Proceedings of the 2012 international symposium on pervasive displays* (pp. 17:1–17:6). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2307798.2307815

Alves, V., Niu, N., Alves, C., & Valença, G. (2010). Requirements engineering for software product lines: A systematic literature review. *Information and Software Technology*, *52*(8), 806–820. Retrieved from https://doi.org/10.1016/j.infsof.2010.03.014

Ardito, C., Buono, P., Costabile, M. F., & Desolda, G. (2015). Interaction with large displays: A survey. *ACM Computing Surveys*, *47*(3). Retrieved from https://doi.org/10.1145/2682623

Bedu, L., Tinh, O., & Petrillo, F. (2019). A tertiary systematic literature review on software visualization. In *2019 working conference on software visualization (vissoft)* (pp. 33–44). Piscataway, NJ, USA: IEEE. Retrieved from https://doi.org/10.1109/VISSOFT.2019.00013

Biehl, J. T., Czerwinski, M., Smith, G., & Robertson, G. G. (2007). Fastdash: A visual dashboard for fostering awareness in software teams. In *Proceedings of the sigchi conference* *on human factors in computing systems* (pp. 1313–1322). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/1240624.1240823

Bjørn, P., & Boulus-Rødje, N. (2015). Studying technologies in practice: "bounding practices" when investigating socially embedded technologies. In V. Wulf, K. Schmidt, & D. Randall (Eds.), *Designing socially embedded technologies in the real-world* (pp. 341–356). London, England: Springer.

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality and Quantity*, *36*(4), 391–409. Retrieved from https://doi.org/10.1023/A:1020909529486

Börner, D., Kalz, M., & Specht, M. (2013). Closer to you: Reviewing the application, design, and evaluation of ambient displays. *International Journal of Ambient Computing and Intelligence*, *5*(3), 16–31. Retrieved from https://doi.org/10.4018/ijaci.2013070102

Brignull, H., & Rogers, Y. (2003). Enticing people to interact with large public displays in public spaces. In *Human-computer interaction interact '03: Ifip tc13 international conference on human-computer interaction* (pp. 17–24). Amsterdam, Netherlands: IOS Press.

Brown, B., Bødker, S., & Höök, K. (2017). Does hci scale? scale hacking and the relevance of hci. *Interactions*, *24*(5), 28–33. Retrieved from https://doi.org/10.1145/3125387

Bryant, A. (2009). Grounded theory and pragmatism: The curious case of anselm strauss. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *10*(3), Art. 2. Retrieved from https://doi.org/10.17169/fqs-10.3.1358

Colley, A., Häkkilä, J., Forsman, M.-T., Pfleging, B., & Alt, F. (2018). Car exterior surface displays: Exploration in a real-world context. In *Proceedings of the 7th acm international symposium on pervasive displays.* New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3205873.3205880

Corbin, J. M., & Strauss, A. L. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Thousand Oaks, CA, USA: SAGE Publications.

Crabtree, A., Hemmings, T., & Rodden, T. (2003). The social construction of displays. In K. O'Hara, M. Perry, E. Churchill, & D. Russell (Eds.), *Public and situated displays: Social and interactional aspects of shared display technologies* (pp. 170–190). Dordrecht, Netherlands: Springer.

Diel, S. (2007). *Software visualization: Visualizing the structure, behaviour, and evolution of software.* Berlin & Heidelberg, Germany: Springer.

Dingsøyr, T., Fægri, T. E., & Itkonen, J. (2014). What is large in large-scale? a taxonomy of scale for agile software development. In *Product-focused software process improvement* (pp. 273–276). Cham, Switzerland: Springer. Retrieved from https://doi.org/10.1007/978-3-319-13835-0_20

Dingsøyr, T., Moe, N. B., Fægri, T. E., & Seim, E. A. (2018). Exploring software development at the very large-scale: A revelatory case study and research agenda for agile method adaptation. *Empirical Software Engineering*, *23*(1), 490–520. Retrieved from https://doi.org/10.1007/s10664-017-9524-2

Du, G., Degbelo, A., & Kray, C. (2017). Public displays for public participation in urban settings: A survey. In *Proceedings of the 6th acm international symposium on pervasive displays* (pp. 1–9). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3078810.3078825

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*, *50*(9–10), 833–859. Retrieved from https://doi.org/10.1016/j.infsof.2008.01.006

Elhart, I., Mikusz, M., Mora, C. G., Langheinrich, M., & Davies, N. (2017). Audience monitor: An open source tool for tracking audience mobility in front of pervasive displays. In *Proceedings of the 6th acm international symposium on pervasive displays* (pp. 1–8). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3078810.3078823

Gallacher, S., O'Connor, J., Bird, J., Rogers, Y., Capra, L., Harrison, D., & Marshall, P. (2015). Mood squeezer: Lightening up the workplace through playful and lightweight interactions.

In *Proceedings of the 18th acm conference on computer supported cooperative work & social computing* (pp. 891–902). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2675133.2675170

Ghare, M., Pafla, M., Wong, C., Wallace, J. R., & Scott, S. D. (2018). Increasing passersby engagement with public large interactive displays: A study of proxemics and conation. In *Proceedings of the 2018 acm international conference on interactive surfaces and spaces* (pp. 19–32). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3279778.3279789

Glaser, B. G. (1978). *Theoretical sensitivity.* Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. (1992). *Emergence versus forcing: Basics of grounded theory analysis.* Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. (1998). *Doing grounded theory: Issues and discussions.* Mill Valley, CA, USA: Sociology Press.

Glaser, B. G. (2008). *Doing quantitative grounded theory.* Mill Valley, CA, USA: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Piscataway, NJ, USA: AldineTransaction.

Grix, J. (2010). *The foundations of research* (2nd ed.). London, England: Palgrave Macmillan.

Gross, T., Stary, C., & Totter, A. (2005). User-centered awareness in computer-supported cooperative work-systems: Structured embedding of findings from social sciences. *International Journal of Human–Computer Interaction*, *18*(3), 323–360. Retrieved from https://doi.org/10.1207/s15327590ijhc1803_5

Gutwin, C., Greenberg, S., & Roseman, M. (1996). Workspace awareness in real-time distributed groupware: Framework, widgets, and evaluation. In *People and computers xi* (pp. 281–298). London, England: Springer. Retrieved from https://doi.org/10.1007/978-1-4471-3588-3_18

Hazlewood, W. R., Stolterman, E., & Connelly, K. (2011). Issues in evaluating ambient displays in the wild: Two case studies. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 877–886). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/1978942.1979071

Huang, E. M., Mynatt, E. D., Russell, D. M., & Sue, A. E. (2006). Secrets to success and fatal flaws: The design of large-display groupware. *IEEE Computer Graphics and Applications*, *26*(1), 37–45. Retrieved from https://doi.org/10.1109/MCG.2006.21

Jakobsen, M. R., Fernandez, R., Czerwinski, M., Inkpen, K., Kulyk, O., & Robertson, G. G. (2009). Wipdash: Work item and people dashboard for software development teams. In *Human-computer interaction – interact 2009* (pp. 791–804). Berlin & Heidelberg, Germany: Springer. Retrieved from https://doi.org/10.1007/978-3-642-03658-3_83

Jurmu, M., Ventä-Olkkonen, L., Lanamäki, A., Kukka, H., Iivari, N., & Kuutti, K. (2016). Emergent practice as a methodological lens for public displays in-the-wild. In *Proceedings of the 5th acm international symposium on pervasive displays* (pp. 124–131). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2914920.2915007

Keskinen, T., Hakulinen, J., Heimonen, T., Turunen, M., Sharma, S., Miettinen, T., & Luhtala, M. (2013). Evaluating the experiential user experience of public display applications in the wild. In *Proceedings of the 12th international conference on mobile and ubiquitous multimedia* (pp. 7:1–7:10). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2541831.2541840

Koch, M., Ott, F., & Richter, A. (2014). Sociotechnically integrated access to virtual communities with communitymirrors. In J. M. Leimeister & B. Rajagopalan (Eds.), *Virtual communities* (pp. 145–161). New York, NY, USA: Routledge.

Koch, M., von Luck, K., Schwarzer, J., & Draheim, S. (2018). The novelty effect in large display deployments – experiences and lessons-learned for evaluating prototypes. In *Proceedings of 16th european conference on computer-supported cooperative work - exploratory papers.* European Society for Socially Embedded Technologies (EUSSET). Retrieved from https://doi.org/10.18420/ecscw2018_3

Mäkelä, V., Heimonen, T., & Turunen, M. (2018). Semi-automated, large-scale evaluation of public displays. *International Journal of Human–Computer Interaction*, *34*(6), 491–505. Retrieved from https://doi.org/10.1080/10447318.2017.1367905

Mäkelä, V., Radiah, R., Alsherif, S., Khamis, M., Xiao, C., Borchert, L., ... Alt, F. (2020). Virtual field studies: Conducting studies on public displays in virtual reality. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–15). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3313831.3376796

Mäkelä, V., Sharma, S., Hakulinen, J., Heimonen, T., & Turunen, M. (2017). Challenges in public display deployments: A taxonomy of external factors. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 3426–3475). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3025453.3025798

Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., & Ames, M. (2003). Heuristic evaluation of ambient displays. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 169–176). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/642611.642642

Markopoulos, P. (2009). A design framework for awareness systems. In P. Markopoulos, B. De Ruyter, & W. Mackay (Eds.), *Awareness systems: Advances in theory, methodology and design* (pp. 49–72). London, England: Springer.

Marquardt, N., & Greenberg, S. (2015). *Proxemic interactions: From theory to practice.* San Rafael, CA, USA: Morgan & Claypool Publishers.

Mattila, A.-L., Ihantola, P., Kilamo, T., Luoto, A., Nurminen, M., & Väätäjä, H. (2016). Software visualization today: Systematic literature review. In *Proceedings of the 20th international academic mindtrek conference* (pp. 262–271). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2994310.2994327

Merino, L., Ghafari, M., Anslow, C., & Nierstrasz, O. (2018). A systematic literature review of software visualization evaluation. *Journal of Systems and Software*, *144*, 165–180. Retrieved from https://doi.org/10.1016/j.jss.2018.06.027

Michelis, D., & Müller, J. (2011). The audience funnel: Observations of gesture based interaction with multiple large displays in a city center. *International Journal of Human–Computer Interaction*, *27*(6), 562–579. Retrieved from https://doi.org/10.1080/10447318.2011.555299

Morse, J. M., Stern, P. N., Corbin, J., Bowers, B., Clarke, A. E., & Charmaz, K. (2009). *Developing grounded theory: The second generation.* Abingdon, England: Routledge.

Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., & Krüger, A. (2009). Display blindness: The effect of expectations on attention towards digital signage. In *Pervasive computing* (pp. 1–8). Berlin & Heidelberg, Germany: Springer. Retrieved from https://doi.org/10.1007/978-3-642-01516-8_1

Muller, M. (2014). Curiosity, creativity, and surprise as analytic tools: Grounded theory method. In J. S. Olson & W. A. Kellogg (Eds.), *Ways of knowing in hci* (pp. 25–48). New York, NY, USA: Springer.

Nord, R. L., Ozkaya, I., & Kruchten, P. (2014). Agile in distress: Architecture to the rescue. In *Agile methods. large-scale development, refactoring, testing, and estimation* (pp. 43–57). Cham, Switzerland: Springer. Retrieved from https://doi.org/10.1007/978-3-319-14358-3_5

Nunamaker, J. F., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The last research mile: Achieving both rigor and relevance in information systems research. *Journal of Management Information Systems*, *32*(3), 10–47. Retrieved from https://doi.org/10.1080/07421222.2015.1094961

Ojala, T., Kostakos, V., Kukka, H., Heikkinen, T., Linden, T., Jurmu, M., ... Zanni, D. (2012). Multipurpose interactive public displays in the wild: Three years later. *Computer*, *45*(5), 42–49. Retrieved from https://doi.org/10.1109/MC.2012.115

Ott, F., & Koch, M. (2012). Social software beyond the desktop – ambient awareness and ubiquitous activity streaming. *it - Information Technology*, *54*(5), 243–252. Retrieved from https://doi.org/10.1524/itit.2012.0687

Paasivaara, M., Lassenius, C., & Heikkilä, V. T. (2012). Inter-team coordination in large-scale globally distributed scrum: Do scrum-of-scrums really work? In *Proceedings of the 2012 acm-ieee international symposium on empirical software engineering and measurement* (pp. 235–238). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2372251.2372294

Paredes, J., Anslow, C., & Maurer, F. (2014). Information visualization for agile software development. In *2014 second ieee working conference on software visualization* (pp. 157–166). Piscataway, NJ, USA: IEEE. Retrieved from https://doi.org/10.1109/VISSOFT.2014.32

Parker, C., Tomitsch, M., Davies, N., Valkanova, N., & Kay, J. (2020). Foundations for designing public interactive displays that provide value to users. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–12). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3313831.3376532

Parker, C., Tomitsch, M., & Kay, J. (2018). Does the public still look at public displays? a field observation of public displays in the wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *2*(2), Art. 73. Retrieved from https://doi.org/10.1145/3214276

Parnin, C., Görg, C., & Rugaber, S. (2010). Codepad: Interactive spaces for maintaining concentration in programming environments. In *Proceedings of the 5th international symposium on software visualization* (pp. 15–24). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/1879211.1879217

Pousman, Z., & Stasko, J. (2006). A taxonomy of ambient information systems: Four patterns of design. In *Proceedings of the working conference on advanced visual interfaces* (pp. 67–74). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/1133265.1133277

Preim, B., Ropinski, T., & Isenberg, P. (2018). A critical analysis of the evaluation practice in medical visualization. In *Eurographics workshop on visual computing for biology and medicine* (pp. 45–56). Geneva, Switzerland: Eurographics Association. Retrieved from https://doi.org/10.2312/vcbm.20181228

Rittenbruch, M. (2016). Cubit: Design and evaluation of a collaboration-tool for large interactive wall surfaces. In C. Anslow, P. Campos, & J. Jorge (Eds.), *Collaboration meets interactive spaces* (pp. 125–152). Cham, Switzerland: Springer.

Rolland, K. H. (2016). Scaling across knowledge boundaries: A case study of a large-scale agile software development project. In *Proceedings of the scientific workshop proceedings of xp2016* (pp. 5:1–5:5). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2962695.2962700

Rubegni, E., Memarovic, N., & Langheinrich, M. (2011). Talking to strangers: Using large public displays to facilitate social interaction. In *Design, user experience, and usability. theory, methods, tools and practice* (pp. 195–204). Berlin & Heidelberg, Germany: Springer. Retrieved from https://doi.org/10.1007/978-3-642-21708-1_23

Schwaber, K., & Beedle, M. (2001). *Agile software development with scrum.* Upper Saddle River, NJ, USA: Prentice Hall.

Schwarzer, J., Draheim, S., von Luck, K., Wang, Q., Casaseca, P., & Grecos, C. (2016). Ambient surfaces: Interactive displays in the informative workspace of co-located scrum teams. In *Proceedings of the 9th nordic conference on human-computer interaction* (pp. 69:1–69:4). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2971485.2971493

Schwarzer, J., Draheim, S., von Luck, K., Wang, Q., & Grecos, C. (2021). *Supplementary material to the ambient surfaces field deployment study (2014–2019).* Mendeley Data, V1. Retrieved from http://dx.doi.org/10.17632/mbzgcrz4zy.1 (Accessed 2 July 2021)

Schwarzer, J., von Luck, K., Draheim, S., & Koch, M. (2019). Towards methodological guidance for longitudinal ambient display in situ research. In *Proceedings of 17th european conference on computer-supported cooperative work - exploratory papers.* European Society for Socially Embedded Technologies (EUSSET). Retrieved from https://doi.org/10.18420/ecscw2019_ep07

Scott-Hill, B., Anslow, C., Ferreira, J., Kropp, M., Mateescu, M., & Meier, A. (2020). Visualizing progress tracking for software teams on large collaborative touch displays. In *2020 ieee symposium on visual languages and human-centric computing (vl/hcc)* (pp. 1–5). Piscataway, NJ, USA: IEEE. Retrieved from https://doi.org/10.1109/VL/HCC50065.2020.9127286

Sharma, V. S., Mehra, R., Kaulgud, V., & Podder, S. (2019). An extended reality approach for creating immersive software project workspaces. In *2019 ieee/acm 12th international workshop on cooperative and human aspects of software engineering (chase)* (pp. 27–30). Piscataway, NJ, USA: IEEE. Retrieved from https://doi.org/10.1109/CHASE.2019.00013

Shelton, B. (2020). *Defining, classifying and evaluating ambient display* (Doctoral dissertation, University of Newcastle, Newcastle, New South Wales, Australia). Retrieved from http://hdl.handle.net/1959.13/1418427

Shelton, B., & Nesbitt, K. (2016). The aesthetic awareness display: A new design pattern for ambient information systems. In *Proceedings of the australasian computer science week multiconference* (pp. 50:1–50:10). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2843043.2843371

Shelton, B., & Nesbitt, K. (2017). Evaluating wavewatch: An ambient display of web traffic. In *Proceedings of the australasian computer science week multiconference* (pp. 9:1–9:9). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/3014812.3014821

Shelton, B., Nesbitt, K., Thorpe, A., & Eidels, A. (2020). Gauging the utility of ambient displays by measuring cognitive load. *Cognition, Technology & Work*, 1–22. Retrieved from https://doi.org/10.1007/s10111-020-00639-8

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 ieee symposium on visual languages* (pp. 336–343). Piscataway, NJ, USA: IEEE. Retrieved from https://doi.org/10.1109/VL.1996.545307

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). Grand challenges for hci researchers. *Interactions*, *23*(5), 24–25. Retrieved from https://doi.org/10.1145/2977645

Siek, K. A., Hayes, G. R., Newman, M. W., & Tang, J. C. (2014). Field deployments: Knowing from using in context. In J. S. Olson & W. A. Kellogg (Eds.), *Ways of knowing in hci* (pp. 119–142). New York, NY, USA: Springer.

Stephanidis, C., Salvendy, G., of the Group Margherita Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., … Zhou, J. (2019). Seven hci grand challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229–1269. Retrieved from https://doi.org/10.1080/10447318.2019.1619259

Stol, K.-J., Ralph, P., & Fitzgerald, B. (2016). Grounded theory in software engineering research: A critical review and guidelines. In *Proceedings of the 38th international conference on software engineering* (pp. 120–131). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2884781.2884833

Torrance, H. (2012). Triangulation, respondent validation, and democratic participation in mixed methods research. *Journal of Mixed Methods Research*, *6*(2), 111–123. Retrieved from https://doi.org/10.1177/1558689812437185

Walsh, I. (2015). Using quantitative data in mixed-design grounded theory studies: An enhanced path to formal grounded theory in information systems. *European Journal of Information Systems*, *24*(5), 531–557. Retrieved from https://doi.org/10.1057/ejis.2014.23

Weiser, M., & Brown, J. S. (1995). *Designing calm technology.* Retrieved from https://calmtech.com/papers/designing-calm-technology.html (Accessed 9 March 2021)

Williamson, J. R., & Williamson, J. (2014). Analysing pedestrian traffic around public displays. In *Proceedings of the international symposium on pervasive displays* (pp. 13–18). New York, NY, USA: ACM. Retrieved from https://doi.org/10.1145/2611009.2611022

Williamson, J. R., & Williamson, J. (2017). Understanding public evaluation: Quantifying experimenter intervention. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 3414–3425). New York, NY, USA: ACM. Retrieved from https://

doi.org/10.1145/3025453.3025598

Ye, E., Ye, X., & Liu, C. (2018). Teamwatch: Visualizing development activities using a 3-d city metaphor to improve conflict detection and team awareness. *PLOS ONE*, *13*(3), 1–27. Retrieved from https://doi.org/10.1371/journal.pone.0193562

Biographies

Jan Schwarzer is a postdoc researcher at CSTI. His primary research interests include ambient displays, machine learning, methodology development, field deployment research, and agile software development. He received his PhD in computer science from UWS. Contact him at jan.schwarzer@haw-hamburg.de.

Susanne Draheim is managing director of the research and transfer institute Smart Systems and the affiliated research laboratories CSTI and Living Place Hamburg. She has a background in Sociology, Educational Sciences and Cultural Sciences and holds a PhD in Sociology from TU Dresden. Contact her at susanne.draheim@haw-hamburg.de.

Kai von Luck is a full professor at HAW Hamburg for computer science and artificial intelligence. He acts as academic director of the research and transfer institute Smart Systems and the affiliated research laboratories CSTI and Living Place Hamburg. Contact him at kai.vonluck@haw-hamburg.de.

Qi Wang is a Professor at UWS. He has served as a Board Member of the EU 5G-PPP Technology Board, Scotland's Developing AI and AI Enabled Products and Services Working Group, and ITU-T Focus Group on Autonomous Networks. He is a winner of UK Times Higher Education (THE) Awards 2020. Contact him at qi.wang@uws.ac.uk.

Christos Grecos (SM IEEE 2006, SM SPIE 2008) is the Chair and Full Professor of the Computer Science Department at Arkansas State University. His research interests include image/video compression standards, image/video processing and analysis, image/video networking and computer vision. Contact him at grecoschristos@gmail.com.