Published by the Open Science Centre, University of Jyväskylä, Finland

ISSN: 1795-6889

HUMAN TECHNOLOGY

Volume 15, Number 1, February 2019

Jukka Jouhki
Editor in Chief

www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

Volume 15, Number 1, February 2019

DOI: http://dx.doi.org/10.17011/ht/urn.201902201604

Contents

Original Articles

HUMAN TECHNOLOGY

ISSN: 1795-6889
humantechnology.jyu.fi

Submission: humantechnologypublishing.jyu.fi

Human Technology is an interdisciplinary, scholarly journal publishing innovative, peer-reviewed articles exploring the issues and challenges within human–technology interaction and the human role in all areas of ICT-infused societies.

Human Technology, published by the Open Science Centre, University of Jyväskylä, is distributed without a charge online.

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 15(1), February 2019, 1–5

From the Editor in Chief

HUMANS AND THEIR TECHNOLOGIES PLAY THE INFINITE GAME

Jukka Jouhki
Department of History and Ethnology
University of Jyväskylä
Finland

The authors of the current *Human Technology* issue approach the relationship between humans and their technologies from quite different angles. Two articles focus on chat technology. In the first, by **Khushnood Naqshbandi** and her associates (Simon Hoermann, David Milne, Dorian Peters, Benjamin Davies, Sophie Potter, & Rafael A. Calvo), the authors focused on codesigning a chat tool for a voluntary-sector organization based on the experiences and needs of the volunteer workers. Meanwhile, in the second, **Marita Skuvje** and her colleagues (Ida Maria Haugstveit, Asbjørn Følstad, & Petter Bae Brandtzaeg) explored the differences between communicating with a chatbot and a chat human, particularly when the chatbot's behavior approaches that of human interaction. **Naveen Kumar and Jyoti Kumar**, in their paper, discussed the next industrial revolution—Industry 4.0—and proposed a framework based on multiple formulae that address the various aspects of both the human and machine components in measuring the efficiency of future smart factories. However, the article by **Subadra Panchanadeswaran** and her coauthors (Ardra Manasi Unnithan, Shubha Chacko, Michael Brazda, Natalie Brooks Wilson, & Santushi Kuruppu) presents a very different perspective of smartness, that of smart phones used by sex workers in India and how this technology has transformed both the work and the workers. Finally, **Patrick Lehane**'s paper on software design explores a theme that is, in some form or another, present in all articles in this issue: the intuitive use of technology.

Technological progress is not simply about inventing applications that automate activities previously conducted by humans. People still play key roles in operating a plethora of technological applications—to tell technologies what to do, where to go, what to produce, who to contact, and so on. No matter how complicated the automated machine, the human always is at some end to use, service, or manage it. However, that human must be trained—or she or he must otherwise feel confident enough in the skills needed—to operate the device. Intuitive use

DOI: https://doi.org/10.17011/ht/urn.201902201605

means that the human interactant feels familiar with an application even if it is new to him or her. This goal of intuitive interaction is emphasized nowadays, particularly during the development of smart consumer applications.

Indeed, many have noticed how the former thick user manuals have shrunk into palm-sized booklets with just a few images and a few lines of mainly inspirational text because the applications are so instinctive to use. Moreover, consumers often do not regard the use of some devices as technology because they have grown so accustomed to them. As Alan Kay, an American computer scientist, has been quoted, technology is “anything that was invented after you were born” (Kay, cited in Larsson, 2011, p. 441). Although Kay’s definition was meant to be humorous, it is true that recent innovations are more likely to attract attention than old technology (e.g., chairs, lamps, cars, TVs) because the latter are viewed merely as objects of normal life. In the same vein, Danny Hillis, a computer scientist and an inventor, defined technology as “everything that doesn’t work yet” (cited in Kelly, 2007).

Intuitiveness leads to efficiency when the use of technology requires little or no need for training and learning; it also makes an application more lucrative to potential users. That is why the development of technology today is increasingly about automation, making equipment work (together) seamlessly, and whatever technology that cannot be automated is made as easy as possible for the human user (e.g., Endsley, 2017; Moore, Upchurch, & Whittaker, 2018, p. 4). This is evident in Kumar and Kumar’s article about future smart factories, the efficiency of which, when measured, takes into account the cognitive load placed upon the human operators. As long as smart factories are not smart enough to take care of themselves, the people they depend on should feel competent and at ease in operating them. Intuitiveness is present also in the two articles about chat technology in this issue. Operating an online chat service between people or people and bots should be smooth and natural. Intuitive interaction is important even—and perhaps most importantly—among the functional illiterate, such as the Indian sex workers who use smart phones to facilitate their everyday work and lives.

Intuitive use often is a step in the evolution of technological innovation between its rudimentary not-at-all-intuitive stage and its final frontier, which could be considered full automation. Achieving full automation will liberate people to devote their energy to something other than the often tiresome, time-consuming, or boring tasks they could gladly do without, no matter how intuitive these tasks are to complete. Just watch the ingenious TedTalk presentation titled *The Magic Washing Machine* by Hans Rosling (2010). In it, Rosling illustrated how washing machines could be considered as “book machines” because they liberated users, particularly women, to educate themselves instead of toiling with some of the time-consuming and menial tasks related to laundry.

One can easily connect Rosling’s idea to mathematician Alfred N. Whitehead’s statement from more than a century ago. He wrote, “Civilization advances by extending the number of important operations which we can perform without thinking of them” (Whitehead, 1911, p. 61). In the views of Rosling and Whitehead, automation takes humanity further toward personal fulfillment. Or, if that sounds too grandiose, the extra time available at least enables people to enjoy something more entertaining and/or allows more comfortability in daily life. Some would be equally justified in saying automation is making people entertain themselves to death by driving them into the mind-numbing autoplay of games, TV series, and films, as well as making everything too easy.

As readers of *Human Technology* know, we editors of this journal emphasize the human side of technology. However, we often contemplate what the human side constitutes, especially for the future. When intuitive use turns into automated, embedded, instinctive, cybernetic, or even genetic use, we humans might end up using technology as we “use” our internal organs today—mainly unaware of their existence. It is likely that future ubiquitous technology will take care of human needs without the recipients even knowing it, just as applications such as automatic air conditioning today are designed to cater to a person’s needs and preferences in a way that makes the user almost oblivious to the technology. Perhaps in the future, the concept of use will mean reading an occasional report created by an artificial intelligence informing any and all users about the many personal technological applications they barely are conscious of using—or perhaps which are *using them*—on the backstage of daily existence. That is, if we were interested in such reports.

Most predictions of artificial intelligence development, such as the one published by the Future of Humanity Institute at the University of Oxford (see, e.g., “Experts predict,” 2017), foresee full automation of labor eventually. In that case, the intriguing human–technology question is, what to do in a society where humans do not need to work? Or are such visions really just a dream (or a nightmare) because work will always be an integral social and moral value to the extent that if work became extinct, it would have to be resurrected (see, e.g., European Group of Ethics in Science and New Technologies, 2018; Schwartz, 2015)?

In a way, the extinction of work started already thousands of years ago, ever since the cultural evolution of early humankind started “sucking orderliness from the environment,” to use Leslie White’s (1959/2016, p. 35) vivid expression, and people began to harness energy with elementary forms of technology to accumulate resources. The resulting surplus of resources enabled specialized professions that were “not really work” because they were not so vital to the physical survival of the society. Rather, they were geared toward inspiring, motivating, and entertaining people. In a way, the priests, jesters, storytellers, and artists of that era (and every era since) could thank technology for their jobs, and their audiences could thank technology for more fulfilling and entertaining lives. Thanks to automation, most of the present-day professions could be said to belong to the same category of not so vital, or perhaps even pointless, as Professor David Graber (2016) claims.

Different views of the social effects of total automation range from moral condemnation to hedonistic utopianism (e.g., Brynjolfsson & McAfee, 2014; Marcuse, 1964/2007). But no matter which view one might hold, it is clear that technology development presents an “infinite game” that is not played to win, but instead just to keep on playing (Kelly, 2007; see also Carse, 1986). Development of technology most likely will never end because new forms of automation liberate humans to think of other new forms. In this respect, studying automation, robotics, artificial intelligence, and the future of technology in general should not aim any lower than asking what is the value of work, if not the meaning of life. When technology replaces most human work, should individuals panic or celebrate, like in the two scenarios depicted in the famous meme in Figure 1? The answer depends on a plethora of complex and frustratingly divisive cultural, social, and political questions revolving around the relationship between an individual and the society.

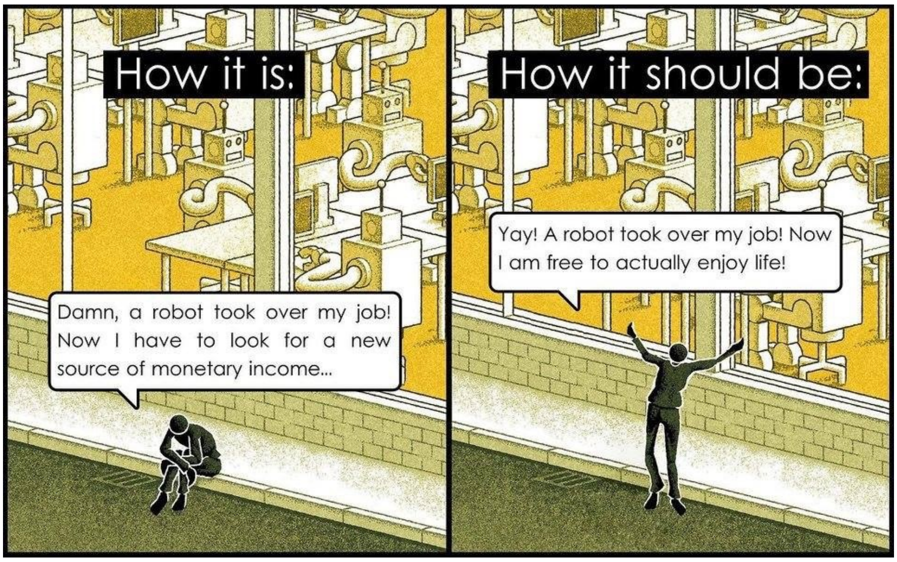


Figure 1. As technology is taking over human jobs, should the reaction be to shout out "The end is near!" or "The beginning is near!" (Picture source: Trox, 2016.)

Fortunately, there are forums like *Human Technology* in which such questions are welcomed to be contemplated. So, if you feel like tackling the issue, feel free to submit an article or even suggest a themed issue, and we can explore the fascinating future of technology more.

REFERENCES

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* London, UK: W. W. Norton & Company.

Carse, J. P. (1986). *Finite and infinite games.* New York, NY, USA: The Free Press.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors*, *59*(1), 5–27.

Experts predict when artificial intelligence will exceed human performance. (2017, May 31). *MIT Technology Review*. Retrieved from www.technologyreview.com/s/607970/experts-predict-when-artificial-intelligence-will-exceed-human-performance/

European Group of Ethics in Science and New Technologies. (2018). *Future of work, future of society* [Opinion No. 30]. Brussels, Belgium: Directorate-General for Research and Innovation, European Commission. Retrieved from https://ec.europa.eu/info/sites/info/files/research_and_innovation/ege/ege_future-of-work_opinion_122018.pdf

Graeber, D. (2016, September 27). Why capitalism creates pointless jobs. *Evonomics: The next evolution of economics.* Retrieved from http://evonomics.com/why-capitalism-creates-pointless-jobs-david-graeber

Kelly, K. (2007, February 22). Everything that doesn't work yet. *The Technium* [Web log]. Retrieved from https://kk.org/thetechnium/everything-that

Larsson, M. (2011). Dune: Arenaceous anti-desertification architecture. In V. Badescu & R. B. Cathcart (Eds.), *Macro-engineering seawater in unique environments. Arid Lowlands and Water Bodies Rehabilitation* (pp. 431–464). New York, NY, USA: Springer.

Marcuse, H. (2007). *One-dimensional man: Studies in the ideology of advanced industrial society.* New York, NY, USA: Routledge. (Original work published in 1964)

Moore, P. V., Upchurch, M., & Whittaker, X. (2018). Humans are machines at work: Monitoring, surveillance and automation in contemporary capitalism. In P. V. Moore, M. Upchurch, & X. Whittaker (Eds.), *Humans and machines at work: Monitoring, surveillance and automation in contemporary capitalism* (pp. 1–16). London, UK: Palgrave.

Rosling, H. (2010, December). *Hans Rosling: The magic washing machine* [Video file]. Retrieved from www.ted.com/talks/hans_rosling_and_the_magic_washing_machine

Schwartz, B. (2015). *Why we work.* New York, NY, USA: Simon & Schuster.

Trox, N. (2016, November 30). [A meme posted in *Basic Income* group of Google +]. Retrieved from https://plus.google.com/101203173543292595997/posts/CwX8vTShPMK

White, L. A. (2016). *The evolution of culture: The development of civilization to the fall of Rome.* New York, USA: Routledge. (Original work published 1959)

Whitehead, A. N. (1911). *An introduction to mathematics.* New York, NY, USA: Henry Holt and Company.

Author's Note

All correspondence should be addressed to
Jukka Jouhki
University of Jyväskylä
P.O. Box 35
40014 University of Jyväskylä, FINLAND
jukka.jouhki@jyu.fi

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

Human Technology

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 15(1), February 2019, 6–29

CODESIGNING TECHNOLOGY FOR A VOLUNTARY-SECTOR ORGANIZATION

Khushnood Naqshbandi
Wellbeing Technology Laboratory
The University of Sydney
Australia

Simon Hoermann
HIT Laboratory
University of Canterbury
Christchurch, New Zealand

David Milne
Wellbeing Technology Laboratory
The University of Sydney
Australia

Dorian Peters
Wellbeing Technology Laboratory
The University of Sydney
Australia

Benjamin Davies
ReachOut Australia
Sydney, Australia

Sophie Potter
ReachOut Australia
Sydney, Australia

Rafael A. Calvo
Wellbeing Technology Laboratory
The University of Sydney
Australia

Abstract: *This paper presents an investigation into the experiences and perceptions of volunteers and community managers of an Australian voluntary-sector organization that supports young help-seeking people. The process focused specifically on the design of a chat tool, a rudimentary version of which was conceptualized and tested during a trial completed prior to this study. The process explored the motivations and experiences of these volunteers using a codesign approach, which led to the development of specific features of the chat tool that were tailored to the nature of their work and organization, as well as the sector-specific ethos. We employed several research methods, which included interviews, focus groups, and participatory design workshops. Thematic analyses were performed on the resultant qualitative data. The methods, motivational themes, and the ensuing design solutions that were implemented are discussed in detail with the aim of encouraging codesign of technology for voluntary-sector organizations.*

Keywords: *chat system, codesign, participatory design, volunteers, human–computer interaction.*

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

DOI: https:/doi.org/10.17011/ht/urn.201902201606

INTRODUCTION

The design of any technology requires a deep understanding of the users and the nature of work of the users (Rogers, Sharp, & Preece, 2011). While the research on technology in private and for-profit sectors has made significant advancements in the last few decades, research on technology in the voluntary sector[1] remains dismal, with a few initiatives undertaken sporadically (McPhail, Costantino, Bruckmann, Barclay, & Clement, 1998; Morse, Cerretani, Halai, Laing, & Perez, 2008; Schümmer & Haake, 2010). A plausible reason for this could be that many routine tasks performed in voluntary-sector organizations are similar to those in for-profit organizations, so naturally, one could assume that technology used in the latter can be successfully adapted for the former. For instance, an integrated customer chat service like Intercom[2] can be used both by a bank employee and by a volunteer working for the Red Cross. Thus, most systems are designed for the for-profit sector and nonprofits generally follow suit (Balser, 2008). Another reason to consider is the lack of capital for investing in organizational technology in the voluntary sector, which makes this sector rely heavily on private donations and grants from the governments in developed nations (Balser, 2008).

Research has shown that the voluntary sector differs from the for-profit sector in terms of work motivation and demographics of its workers (Leete, 2000). It is worth pointing out that this sector uses the service of volunteers abundantly, individuals who are not usually driven by explicit material considerations (Bussell & Forbes, 2002). The importance of customizing technology to the voluntary context and its people has been recognized as the key to the success of these organizations (Balser, 2008). Moreover, technology changes work processes, tasks, job satisfaction, workload, and power relationships in the voluntary sector (Saidel & Cour, 2003). Thus, it is essential to emphasize the importance of creating a bridge between the voluntary sector and human–computer interaction (HCI) research so that the latter better addresses the need for designing technology for this sector.

Establishing requirements for an interactive system requires elicitation of those requirements using suitable methods (Rogers et al., 2011). In the past, the field of software engineering has employed some very rigorous requirements-gathering methods, typically focusing on accuracy and quantitative measurement in software development, which reflects its conventional engineering roots (Preece & Rombach, 1994). The gradual shift of focus to usability and the user in a holistic perspective has resulted in software engineering adopting more flexible and creative approaches toward designing and developing systems (Löwgren, 1995). For instance, the core values of the currently popular agile software development methodology reflect this adaptation, as enshrined in the agile manifesto: (a) individuals and interactions over processes and tools; (b) working software over comprehensive documentation; (c) customer collaboration over contract negotiation; and (d) responding to change over following a plan.[3] This and many other methods and techniques owe much to design approaches such as participatory design (Kautz, 2010) and other sociotechnical perspectives that relate the users to technology within specific contexts (Cooper & Foster, 1971).

Codesign is a creative approach used to inspire the design process (Guerrini, 2011). It focuses on the collaboration among the users, designers, and researchers (Sanders & Stappers, 2008). This approach also is known as participatory design, which has roots in the Scandinavian labor movement that emphasized involving the workers in the design of their work environment (Schuler & Namioka, 1993). The theory behind this approach is that the tacit knowledge of tasks cannot completely be formalized or quantified but rather needs to follow a flexible process of codesign and coresearch (Spinuzzi, 2005). Depending on the discipline, context, or research or design stage—and numerous other factors—the codesign approach usually involves creative techniques such as interviewing and group discussions, brainstorming, storyboarding, affinity diagraming, mockups, role playing, and storytelling (Sanders, Brandt, & Binder, 2010; Steen, Manschot, & De Koning, 2011). Such activities generate information-rich qualitative data, which, until recently, were underutilized in the social and behavioral aspects of technology research, given the earlier focus on quantifying data in technical fields (Dybå, Prikladnicki, Rönkkö, Seaman, & Sillito, 2011).

Technology designers have applied the codesign approach in devising a wide range of products and services, and its relevance is highly apparent in sectors that are resource scarce and public-service oriented (Pilemalm, 2018). Studies have shown that involving the users in the design process results in the creation of products and services that are significantly more useful and valuable for the users as well as their participating organizations (Kristensson, Gustafsson, & Archer, 2004; Steen et al., 2011). With respect to technology design in a social context, specifically software and information systems, the codesign approach has been successfully used to design for e-governance (Anthopoulos, Siozos, & Tsoukalas, 2007), education (Penuel, Roschelle, & Shechtman, 2007), nonprofits (McPhail et al., 1998), and community engagement (Merkel et al., 2004), among many others. It is within this context of designing software for a voluntary-sector organization offering public service that we place our project.

BACKGROUND

Digital Peer-Support Service for Youth Mental Health

Providing mental health support to young people is challenging due to the stigma and the debilitating nature of mental illness and the common hesitancy to seek help (Gulliver, Griffiths, & Christensen, 2010). Indeed, only 29% of young people (aged 16–34) in Australia with mental health problems reported using any support services, compared to 40% in older age groups (Australian Bureau of Statistics, 2010). Several studies conducted in other countries representing a wide range of healthcare practices consistently show a similar trend (Clement et al., 2015; Gulliver et al., 2010). This makes an online, specifically online chat, service a quite useful, instantaneous, and cost- and time-effective way of delivering appropriate help while simultaneously maintaining the privacy of the help seeker (Dowling & Rickwood, 2013; Reyes-Portillo et al., 2014; Hoermann, McCabe, Milne, & Calvo, 2017). Peer-support technologies for the provision of mental health services take a step back and, instead of dealing with problems after they become aggravated, they deal with early signs of mental distress and, thus, are proactive rather than reactive. Peer support is an important aspect of a holistic solution to youth mental health support (Fo & O'Donnell, 1974; McGorry, 2007). Instead of relying completely on the clinicians and professional therapists, the peer-support model utilizes trained peers who motivate and gently nudge the mentally distressed youth towards the right mental health and well-being choices. These peers are not professional mental health workers but, in most cases, volunteers who take an active interest in the cause that they associate with (Davidson, Chinman, Sells, & Rowe, 2006).

ReachOut

This study focuses specifically on the design of a peer-to-peer chat system for ReachOut, a voluntary-sector organization dedicated to providing mental health and general well-being support to Australians aged 14–25 years. The aims of ReachOut are early intervention, prevention, and information (ReachOut, 2015). Accordingly, the volunteers do not offer online counseling or psychotherapy, but the outcomes of their peer support for the distressed youth are no less important. The service provides either immediate relief to the young person or helps him/her recognize the need for clinical help. Currently, most online peer-to-peer support in Australia occurs via the ReachOut forum, where trained volunteer peers communicate asynchronously online with help-seeking youth and offer empathetic understanding, personal encouragement, and careful, tailored referral to the appropriate resources.

The proposed ReachOut online chat forum was intended to be workable for the users of the system. These users comprise two main categories: visitors—young people going through tough times and seeking help—and moderators, the online volunteers who are trained by ReachOut to offer peer support to distressed youth (Naqshbandi et al., 2016). The moderators are young adults, aged 16–25 Australian residents, who are recruited and trained by ReachOut to moderate its online discussion forum. These moderators are supervised by and report to professional senior community managers within ReachOut, individuals who have specialized qualifications and experience. At the time of the study, ReachOut had 19 moderators dispersed across Australia who, collectively, had moderated more than 35,000 forum posts from approximately 4,000 visitors within the years 2014–2015 (ReachOut, 2015). These moderators have a good understanding of online technology and feel confident enough to use it for their volunteer work. They generally have some vocational or other volunteer experience or tertiary-level educational qualification in mental health, community affairs, social work, or working with young people. The moderators are categorized as junior or senior moderators based on the regularity and length of their moderation activities. In other words, moderators possessing less than a year of regular experience are junior moderators while those with longer service are senior moderators.

Chat Trial

In order to gather some evidence regarding end-user interest in live chat option, a real-time peer-to-peer chat-based support system was introduced at ReachOut for the first time during a trial that was a precursor to this study (Milne, Hoermann, & Calvo, 2016). This trial introduced a simple version of a text-based chat service to the ReachOut Web site and had pre- and postchat surveys for 84 visitors and the ReachOut moderators. Of the 84 visitors, 21 visitors were deemed as trolls and 18 visitors were out of the eligible age bracket during the prechat survey. Of the 45 remaining visitors, only 29 gave their informed consent. Both the visitors and the moderators rated the trial quite positively—26 out of 29 (90%) participating moderators rated the moderator–visitor chat session as worthwhile, while 18 out of the 29 (62%) visitors who finished the postchat survey thought the same. Ten of the 29 participating visitors did not finish the postchat survey, and one visitor did not think it is worthwhile because he wanted help with homework. Of the visitors who completed the survey, the ones that had reported sad or anxious moods at the start reported feeling better after the chat. The helpful text links shared by the moderators in the chat resulted in a good click rate of 73%.

However, analysis showed that a considerable amount of the moderators' time—even on chat conversations with a visitor who moderators deemed worthwhile—was spent idle, waiting for the visitor to respond (56%), looking for relevant resources (13%), assessing eligibility (14%), and so on. This meant that, at any given time, the moderators could have handled more visitors or made better use of their time and efforts through other services. The busiest hour on ReachOut forum in the year 2015 involved almost 800 visitors, indicating that the moderators often engage multiple help seekers simultaneously. Given the limited number of moderators working actively in the forum at any given time, introducing certain automations could potentially help facilitate the chat moderation process. However, it was not clear from this trial what the moderators would consider appropriate and useful, or what system or service would be in line with theirs and the organization's values. Moreover, the researchers recognized the possibility that other solutions for making ReachOut's online peer-to-peer service more efficient and easy might not have been considered because of the limited nature of the trial. While the behavior of visitors to the ReachOut site has been studied extensively (e.g., Burns, Morey, Lagelée, Mackenzie, & Nicholas, 2007; Collin et al., 2011), the behavior of moderators has not attracted much attention. Consequently, further research was needed to examine the moderators' perspectives and explicitly explore these questions.

A few challenges for engaging with the moderators were known prior to the study. For example, they require great flexibility in deciding how much personal time they can spend on moderating. They are geographically dispersed within Australia and within multiple time zones, thus potentially interfering with communication with one another. More importantly, the community managers, who are paid employees of the organization, felt that it was essential to make sure that the moderators were well supported throughout the emotionally demanding task of providing synchronous chat support to vulnerable youth. Additionally, the dynamic between the moderators and their community managers represented an important aspect to be studied. These community managers are responsible for overseeing the moderator recruitment, training, and monitoring, and they serve as the moderators' regular point of contact with the organization and thus, form an essential component for the moderator engagement. The importance of involving community managers for designing volunteer tasks and engagement with an organization is backed by ample evidence in peer-reviewed literature for both traditional and online forms of volunteering (Alfes & Langner, 2017; Eveleigh, Jennett, Blandford, Brohan, & Cox, 2014; Shin & Kleiner, 2003).

Thus, it seemed worthwhile to continue the study in order to propose solutions supported by further research. This research would require a dynamic approach that allowed the researchers to explore redesign of the chat system while considering the multiple social, personal and organizational factors that could impact effectiveness of the system and the engagement of the moderators. Consequently, the research team chose a codesign approach. The goals of the research were (a) understanding moderators' expectations and motivations, (b) determining moderators' current aptitude in using software systems as part of their work, (c) understanding moderators' perceptions of the integration of automation in the chat system, and (d) exploring the dynamic between the moderators and community managers.

METHODOLOGY

This study used a codesign approach that involved the researcher-designers[4] (Authors 1-4, 7), and participants from ReachOut (the community managers, Authors 5 & 6), and the moderators.

The researcher-designers proposed activities with the intention of involving all these parties during the design process. In their planning, the researcher-designers understood the importance of choosing methods that would be most effective toward achieving our shared research goals. This study comprised three stages: an initial workshop with some participatory activities, interviews with individual participants, and a final follow-up workshop with participatory activities. All the interview and workshop sessions were audio recorded for subsequent analyses. For coding practices, we used the general inductive approach (Thomas, 2006).

We carried out this study in accord with the recommendations of University of Sydney ethics conventions. The protocol was approved by the University of Sydney Ethics Committee (Project No. 2016/06). All subjects signed written informed consent agreements in accord with the National Statement on Ethical Conduct in Human Research (2007).

Study Design

To facilitate design considerations, each workshop involved a general focus group discussion around the motivations, goals, and experiences of the moderators, followed by several individual and group activities designed to gather specific system-design requirements. The first workshop session introduced the concept of a peer-to-peer chat system by demonstrating a series of low-fidelity prototypes in order to elicit responses and critiques from the moderators. The subsequent online individual interviews complemented the information gained from the first workshop session. Finally, in the second workshop session, we demonstrated a functional chat system to confirm that the design elements were in line with the needs of the moderators and context of the chats. We conducted several chat simulations in which moderators, community managers, and researcher-designers role-played the moderator–visitor chat sessions, based on the visitor personas extracted from the ReachOut forum, in order to test the proposed concept (Workshop 1) and system (Workshop 2).

The workshops were conducted within the ReachOut headquarters. This location is the usual work facility of the ReachOut community managers. The moderators joined remotely from different parts of the country via Skype. The participant information sheet, consent sheet, workshop materials and artifacts with remote participants were exchanged via email.

Participants

Our collaborators within ReachOut Australia recruited a representative sample of volunteer moderators; all recruited moderators and community managers agreed to voluntary participation. The participating community managers had several years of experience dealing with the moderators in ReachOut. The moderators who participated in the workshops had been actively moderating the online forum for at least a year and were therefore senior moderators. The moderators who participated in the interviews (Stage 2) had less than a year of moderating experience, and thus were junior moderators.

During the course of discussions with all of the participating moderators, we found that most of them were either studying or had recently completed a tertiary-level program in psychology and were interested in using this experience as a means to advance their future career prospects. The community managers were professionals in their line of work and had many years of experience in community work especially related to managing youth mental health. The

ReachOut participants (for all the workshops and the interviews) comprised four female and one male moderators and one male and one female community managers. The moderators were aged in their early 20s; the community managers were in their early- to mid-30s.

The researcher-designers initiated and facilitated the discussions and activities in these workshops as well as the interviews. During the focus group and interview discussions, the researcher-designers encouraged the ReachOut participants to elaborate, elucidate, or reiterate their discussion points so that the researcher-designers could understand the lived experiences of the moderators and community managers. Other, more hands-on activities engaged the researcher-designers and ReachOut participants performing similar or balancing roles. The recommended participant group size—tiny (2-4) to small (6-8), as recommended by Muller and Kuhn (1993)—was reached for both workshops to match their participatory design goals.

Initial Workshop

The initial workshop was conducted with focus group discussions interspersed with creative participatory activities that were designed to elicit responses and reveal underlying motives of the intended end users. These activities included (a) brainstorming, (b) affinity diagramming, (c) sketching, and (d) role-playing in the online chats. This workshop involved four ReachOut participants—two community managers employed full-time by ReachOut and two experienced moderators (who participated remotely via Skype), and four researcher-designers. The workshop lasted for 2 hours. Activity worksheets were emailed to the remote participants, who returned them through email after the activity sessions.

The session began with an informal introduction and ice-breaking session for the first few minutes. We followed it with a semistructured discussion of the moderators' motivations and experiences of volunteering with ReachOut. The discussion was cued in the form of open-ended questions drawn from some of the areas of research interest (see end of section Chat Trial), such as "What do you do as a moderator?," "Why do you work as a moderator?," and "What kind of forum activities are you doing?" The discussions flowed organically most of the time, but were controlled using discussion cues by the researcher-designers if the participants deviated too far from the topic or if the process became prolonged.

The initial discussion was followed by a simulated chat with a moderator using an off-the-shelf chat tool in which one of the researcher-designers adopted the persona of a potential visitor (e.g., a young person seeking information on drug abuse). The moderators worked together to maintain the conversation while openly discussing the visitor's likely motivations and needs and the rationale behind each dialogue move.

The session then moved on to brainstorming and affinity diagramming that involved structuring ideas using Post-it Notes. The ReachOut participants were asked to think about and discuss situations that would be concerning in an online chat conversation and, conversely, to envision an ideal conversation. They also were asked to describe a situation or visitor that would be difficult to deal with and, conversely, a situation or visitor for which they could imagine the conversation going particularly smoothly. The responses were jotted onto Post-it Notes as separate points and gradually structured by all the participants into four topics: concerns, ideal, difficult, and straightforward (see left portion of Figure 1).

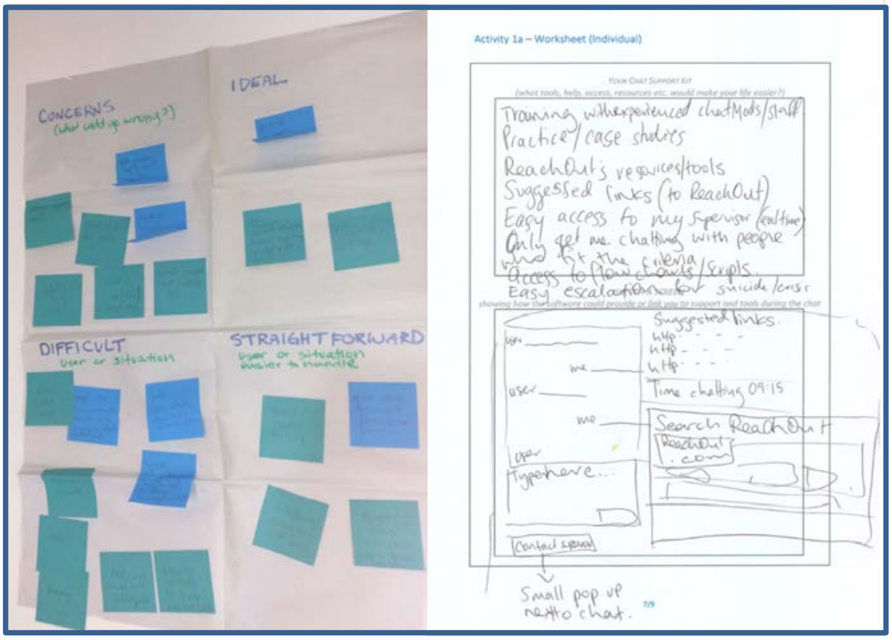


Figure 1. Affinity diagramming (left) and sketching (right) activities in which ReachOut moderators, community managers, and researcher-designers participated during Workshop 1.

Participants then were asked to sketch a "support kit" of useful resources for the chat system. The goal was to generate additional creative insights into the moderation work based on the understanding of the moderators (see right portion of Figure 1).

Two additional simulated chats were conducted via the same protocol described previously. In the first chat, we used the persona was of a concerned youth worried about her friend's drug abuse (see Figure 2), the second of a severely depressed visitor at risk of harm.

Although the primary objective of the workshop was the design of the chat tool, the researcher-designers kept open minds toward other possible digital solutions during the discussions and creative activities. In particular, they avoided presenting any leading questions or cues.

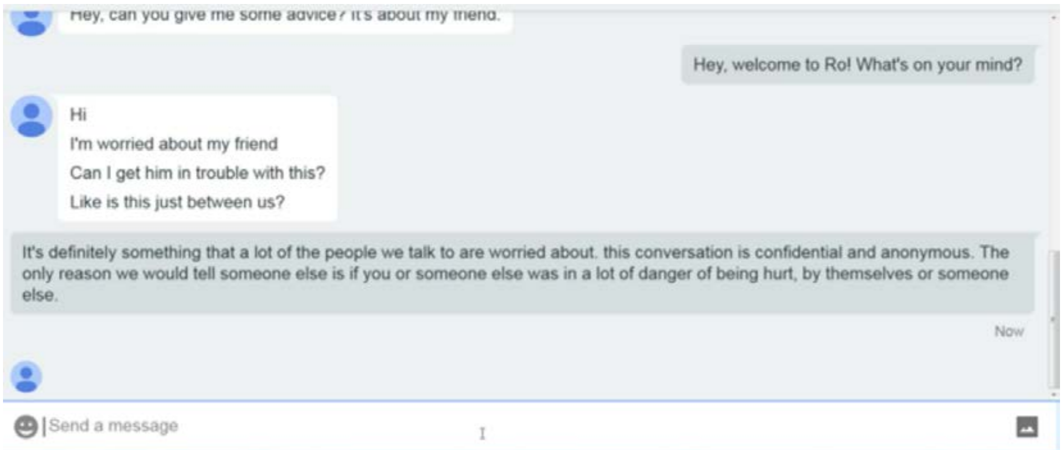


Figure 2. Screenshot of the simulated chat during Workshop 1, where ReachOut moderators and researcher-designers role-played visitor–moderator interactions.

Individual Interviews

Individual interviews were conducted with two moderators, each with less than a year of volunteer experience at ReachOut. These interviews were conducted remotely via Skype, each with one researcher-designer and one moderator. The discussions lasted less than an hour—55 minutes and 45 minutes for Interviewees 1 and 2, respectively.

The interviews were semistructured. Although the researcher-designers provided discussion prompts related to a series of topics of interest, the moderators were encouraged to add to the discussion wherever relevant and were asked probing questions whenever needed. The questions were similar to the ones asked in the focus groups from Workshop 1. Each interview began with a discussion of the participant’s motivations for volunteering and his/her experiences. It then moved to simulated chat sessions in which a researcher-designer adopted the persona of a potential visitor and communicated exclusively via the text-based features of Skype. The participant was encouraged to respond via text while explaining his/her thought process via audio. Later in the interview sessions, the moderators were asked open-ended questions to discuss concerning, difficult, straightforward, and ideal situations and visitors.

Follow-up Workshop

A second workshop was conducted with three participants—one community manager and two moderators (one moderator and the community manager also participated in the first workshop). The purpose of this workshop focused more specifically on the design and features of the potential chat system. The discussion questions were the same as the first workshop to provide a quick recapitulation of earlier points but were time-controlled in order to provide more time for the prototype discussion. The discussions also revisited the topic points—concerning, difficult, straightforward, and ideal situation/visitors—although this time the points were structured categories based on the previous workshop findings. This workshop repeated the simulated chats as the previous workshop but, instead of using an off-the-shelf tool, the participants used a working prototype of the system we will describe in later sections of this paper. Also, rather than researcher-designers alone adopting the visitor personas, the participants themselves alternated between the visitor and moderator roles using personas that were assigned to them by the researcher-designers.

After a firsthand experience with the high-fidelity prototype, the moderators and community managers then discussed the various features of the prototype, as well as its perceived usefulness and whether anything could be improved. For instance, one of the moderators used the term “sticky note” to describe how an ideal tool would help him compartmentalize his thoughts while moderating. The researcher-designers used this and other insights from this workshop to further refine the prototype. The Workshop 2 lasted for 2 hours. Table 1 consolidates some of the information gained from the workshops and interviews.

Analysis

The primary method of collecting data was through audio recording. One of the researcher-designers present during all events transcribed the audio data from the interviews and focus groups/workshops.

Table 1. Summarized Details About the Number of Participants, Overall Duration, and Methods and Activities Associated with the Data Collection.

	ReachOut Participants	Researcher-designers	Overall duration	Methods/ activities
Workshop 1	2 senior moderators 2 community managers	4 researcher-designers	2 hours	Semistructured focus group discussions Brainstorming and affinity diagraming Sketching Role-playing in chat simulation
One-on-one interviews	2 (separate sessions) with junior moderators	1 researcher-designer for each session (also present in both workshops)	Interview 1: 55 minutes, Interview 2: 45 minutes	Semistructured interview discussions Role-playing in chat simulation
Workshop 2	2 experienced moderators (1 participated in Workshop 1), 1 community manager (common with workshop 1)	3 researcher-designers (2 participated in Workshop 1)	2 hours	Semistructured focus group discussions Brainstorming Sketching Role-playing in chat simulation

Transcripts from the interviews and the focus groups/workshops were analyzed using the general inductive approach (Thomas, 2006). This approach uses inductive analysis, where the readings of the raw qualitative data are used to derive themes or concepts. The goal is to avoid any structured constraints influenced by predetermined objectives. Two researcher-designers (first two authors) closely read the texts of the transcripts (including one who transcribed the recordings), independently analyzing the transcripts. During the readings of the transcripts, each researcher-designer identified specific themes, with the aim of capturing the primary messages that the participants conveyed. Consideration was given to the following while outlining the themes: frequency (how often the concept was mentioned), intensity (strength of the comment), specificity (detail of the response), and participant perception of the importance of the concept (Krueger, 1997). An example of a prominent theme is “Building rapport with the visitors.” Some examples of quotes that indicated building rapport with the visitors are *“Trying to ask a lot of questions to get them talking to build up that rapport”* (Moderator 3), and *“A good chat is when you are really engaging with the user. You form a connection with them based on what they have said to you”* (Moderator 4).

During the course of combining the two independent sets of themes, the researcher-designers noticed that many themes expressed similar ideas. For instance, the theme “Following community guidelines” represented comments of this nature,

> *So, the user is in a highly emotive state; they might be prone to say things that are inappropriate or just in the heat of the moment. And if that is the case, you are unable to offer them full support. You have to outline what the guidelines are and inform them of that.”* (Community Manager 1)

Meanwhile, the theme “Getting help if required” included comments such as, “*This is the point where someone else should come in and support the chat*” (Community Manager 1). Because of the similarity in suggesting the same outcome for the moderator, these related comments were combined into a new theme, called “Providing support and guidance to the moderator.” On the other hand, the theme “Promote use of the forum” contained only one coded comment, from Moderator 2, “*You’re also heaps welcome to come and have a chat to us on the forums if you need anything else. We’re here for you. forums.reachout.com,*” and was thus discarded.

The process of coding, identifying, and combining themes was repeated a total of three times, when agreement between the coders was achieved. We discuss the resultant themes in the next section.

RESULTS

In this section, we describe the themes that emerged during qualitative analysis. We categorized the themes into (a) those that relate to participants’ current skills and experience and whether they feel equipped to provide real-time chat support; (b) those that relate to why they would be motivated to volunteer; and (c) the specific details of what the service should try to do and how.

Confidence and Concerns

Participants were generally optimistic and confident about potentially assisting help-seekers via real-time chat. In many ways, they saw it as a natural extension of their existing duties moderating the ReachOut forum. They felt confident that their previous experiences had given them knowledge of and access to a large amount of useful information to share. This included reference materials hosted on ReachOut.com and elsewhere, past conversations posted in the forum, and external services (e.g., HeadSpace, Kids Helpline) for referrals to visitors when appropriate. They considered themselves as just one component in the overall holistic framework for treating these troubled youths.

> *But then there is the community outside as well: family members, parents and other supports, other specialists, crisis services. We are often the beginning of a much bigger, longer journey with a lot more sort of variation, a lot more types of support. We are part of the journey and we push it along in the right direction.* (Moderator 4)

As peers, they saw themselves as uniquely well equipped to offer support with empathy and encouragement. “*The support that peers can give to each other: That is quite separate and different and unique to the support* [you can get] *from self-help or professionals*” (Moderator 2).

However, the participants also expressed concern that the chat service would be more likely to place them outside of their areas of expertise. In the forum, they have a great deal of control over which topics and users they engage. However, in a chat service, they might be assigned to a visitor for whom they have limited capacity to relate to, and conversations might quickly traverse into areas of which they have little knowledge or experience.

Motivations for Participating

When asked about their motivations for volunteering at ReachOut forum, some moderators mentioned motivations related to the greater good, such as making a difference to peoples’ lives and contributing positively to society. The respondents stated that they perceived ReachOut as a safe place in which they could make meaningful contributions by facilitating healthy discussions around the various issues that affect young people. These involve *"building capacities in our community of young people around understanding, recognizing, and supporting"* (Community Manager 2). They also appreciated being part of the online community of peers: "*Yeah, I really enjoy working for ReachOut. It feels like home*" (Moderator 3).

In addition, several participants saw ReachOut as a stepping-stone toward a career in a related field. Disciplines such as psychology and community work came up quite often when discussing their motivation: "*I come from a psychology background. I was looking to get experience within the field of psychology in order to help me to progress further in my career"* (Moderator 1).

Goals and Priorities for the Service

Participants had some very specific feedback about how the chat service should be conducted. We provide these details in the following subsections.

Build Rapport with Visitors

One of the most prominent themes to emerge during the analysis was the need to build a strong rapport with each visitor. This was seen as particularly important during the early phases of a chat conversation, to help visitors feel at ease: "*We don't want to just straight jump into it* [the problem] *either. Like, this is our first interaction with this person, so how are we going to make them welcome?"* (Community manager 1).

Participants predicted that some visitors would have difficulty opening up. Additionally, in some cases, the cause of distress would be ambiguous, and visitors would need help navigating through their emotions, such as *"Someone who is not really self-aware and unable to tell what's going on but obviously is really struggling. So, the challenge is to ask the right type of questions and build the type of rapport"* (Moderator 3).

Uphold Visitor Autonomy

Participants felt strongly averse to prescribing direct solutions to visitors. Instead, they felt the best approach was to empower visitors to formulate their own plans and strategies: "[Our job is to] *help someone make some decision on what they should do next, rather than offer up solutions exactly ourselves"* (Moderator 4).

This autonomy-supportive approach was demonstrated during the simulated chats. The participants were careful to avoid prescriptive statements and wary of jumping to a solution too quickly*: "I am careful to not give direct advice; I guess we are not exactly trained to directly counsel our users. So,* [I] *don't want to exactly tell what the person what to do"* (Moderator 2).

When it felt appropriate during the interaction, moderators would share resources that they felt might be relevant and useful but would generally offer multiple options and avoid making any decisions on the visitor's behalf. "*We have a bunch of resources on our site, so if it's cool, I can give you a few links, but we can also keep chatting too!"* (Moderator 1). *"Which ones do you want to pick? How do you want to take them?"* (Moderator 3).

A community manager explained that this focus on autonomy is a core guideline for the organization, which aims to follow a strengths-based approach to working with young people. *"This is where I am using a strengths-based model to get them to think about ways to help themselves"* (Community manager 1).

Preserve Anonymity and Maintain Boundaries

ReachOut's existing peer-support services are provided anonymously as part of their commitment to providing a safe place for young people to share their private struggles. New users are warned not to enter any identifying information when they create their public profiles, and to not reveal their own (or anyone else's) identity when posting messages on the forum. One of the moderators' core responsibilities is to locate any identifying information on the forum, redact it, and remind users about this policy.

Participants were keen for the chat service to remain anonymous also, given that it would have the same goal of providing a safe place for a young help-seeker to open up. The preference for anonymity extended not only to the visitors' identities, but also to their own. The participants acknowledged that the need for anonymity could have a negative impact on their capacity to build rapport or conduct follow-ups with visitors. A good compromise would be to adopt existing forum user-names and profiles so that visitors would have at least some knowledge with whom they are talking. This would also facilitate making contact via the forum after the chat.

Follow-up with Visitors about Postchat Outcomes

Given the focus on building empathy and rapport during each chat, participants felt it would be natural and desirable to continue checking with them [visitors] from time to time after each chat. *"Follow up questions. Ask them how they are doing, or how their appointment with the psychologist went. Making it a bit more personal for them is a good thing"* (Moderator 1).

They were also keen to know if visitors would follow through with the advice and suggestions they were given. This was seen as valuable information for refining their approach to future conversations, as well as for gaining some validation and positive feedback that these conversations would have a meaningful impact. It would likely be an important component for maintaining long-term motivation and engagement. However, participants also recognized that visitors would be difficult to contact for follow-ups if the service were entirely anonymous. They also felt that it would be important to respect visitors' autonomy and ensure that they had a say in whether any follow-up contact would be made and what form it would take. One simple solution they proposed was to end each chat with an invitation to join the forum and continue the conversation there: *"We're on the forum all the time so come visit"* (Moderator 3).

Respect Moderators' Time and Effort

The participants expressed concern that some conversations or portions of conversations would be a poor use of their time. For example, *"Someone pops up and says 'I wanna chat. How are you guys?' and it's got literally nothing to do with anything and it doesn't go anywhere"* (Moderator 2).

Some specific situations that they considered likely to occur were

- Visitors outside of the target demographic (i.e., Australians aged 14–25) who they, would not be well equipped to help,
- Visitors who intentionally waste time to provoke a reaction (i.e., trolls),
- Visitors who prolong a conversation after it has run its natural course,
- Visitors who do not have anything specific to discuss.

One way to counter these potential situations was to set the right expectations at the start of the conversation: *"…setting the expectations at the beginning so that users do not expect that moderators have infinite time"* (Moderator 4). They also thought it would be useful to have a predefined set of guidelines that could be consulted and cited when visitors engage in undesirable behavior. *"It seems to be like an inappropriate use of the ReachOut service. So,* [I] *need to inform this user of that and outline the guidelines"* (Moderator 3).

A similar document already exists for ReachOut's existing forum and acts as a reference for how users are expected to treat each other. These community guidelines are consulted and cited frequently when dealing with undesirable behavior in the forum and was characterized as mutually beneficial for all involved parties. In addition to that, some participants suggested that they would feel more effective in their work if the system could screen out ineligible visitors, trolls, and those without specific needs.

Provide Guidance and Support for the Moderator

Participants expressed some concern that they have only limited training and will likely encounter situations for which they are not equipped to manage. The moderators would themselves require help on occasion. They recognized that the proposed chat service would be more problematic than the existing forum, where they can choose which posts and users to interact with and can take more time to conduct research and formulate responses.
One reason they might feel out of their depth would be if they encountered a specific circumstance or domain for which they had little training or personal experience to draw from, as Moderator 5 indicated.

> *If someone is in a novel situation that, for whatever reason, you feel not equipped with the proper knowledge about services or other information to provide to the user, then you feel unsure what to do. In such an instance, you want to be able to refer to someone else to get your own sort of assistance.* (Moderator 5)

Distinct from this was the concern that some chats would put them out of their depth simply due to the intensity of the situation. *"Maybe the person is in a really difficult situation; maybe you don't feel equipped to support them in the best way. I guess you need some form of support for the person running the chat as well"* (Moderator 3).

One possible avenue of support that the participants mentioned was to have a reference guide that provided specific steps to follow in various situations. The participants explained that they made frequent use of an existing handbook when moderating the forum. Perhaps this could be adapted for the chat service. *"In a particular area, I would feel less equipped to help someone, but that's what the youth moderator handbook is* [for]. *I followed the directions in there to approaching those kinds of requests"* (Moderator 4).

However, the moderators also recognized that, due to the time constraints imposed by the real-time chat service, this handbook would need to be concise and easy to draw from and/or thoroughly internalized through training and experience. *"Yeah, quick links, kind of like if they are talking about anxiety, depression or whatever, have useful, professional kind of services that you might want to add"* (Moderator 4).

Another avenue mentioned was to provide direct access to a supervisor (i.e., the community manager), who would be able to quickly offer guidance and advice, or even take over particularly difficult or risky situations.

> *So, they might say something like, "It's OK. It won't matter soon." It might be as simple as that. And that is the point where it's probably good for me* [the community manager] *to get involved to help you guys with the risk assessment kind of thing.* (Community manager 1)

Participants also thought it would be useful to be able to offer advice and support for each other—regardless of any seniority. Because they are likely to have different areas of expertise, they could share the workload and stress of having to rapidly compose responses and identify relevant resources.

Find the Right Person for Each Visitor

Participants were concerned that they might not be equipped to help all participants equally. They felt it would be important to assign each visitor to the person best equipped to help them and suggested this could be achieved by allowing moderators (or their supervisor) to specify their level of expertise and the types of conversations that would best suit their expertise. Community manager 1 noted the importance of being able to position moderators in a way that allows them to be successful: *"…being able to assign topics based on skill levels, like 'This person is really new, but they're OK with basic advice from referrals. But they're not OK with crisis and tough time stuff.'"*

The visitor might then be able to choose a specific moderator or be assigned automatically to someone after entering some information about his/her reason for chatting. This information about expertise and topic preference would also be useful for letting moderators know whom to contact if they find themselves out of their depth later on in a conversation.

DISCUSSION AND DESIGN OF THE SYSTEM

Since conducting the workshops, we have designed, developed, and temporarily launched an initial version of a chat system customized specifically to ReachOut's needs. We discuss here the features of this system and how these features reflect the resultant themes and insights drawn from the discussion and activity data.

From the perspective of the visitor, this system differs little from other automated chatbots used in e–commerce and on other Web sites. As shown in Figure 3, it is a small widget that appears on bottom right corner of the ReachOut Web site. The widget will follow the visitor as he/she browses the site, allowing him/her to maintain a conversation and chat about the material encountered.

Figure 4 provides an overview of the system from the perspective of the moderator. The interface is divided horizontally into three main areas: a main menu on the left, the content of the current conversation in the middle, and secondary information or features related to the current conversation on the right.

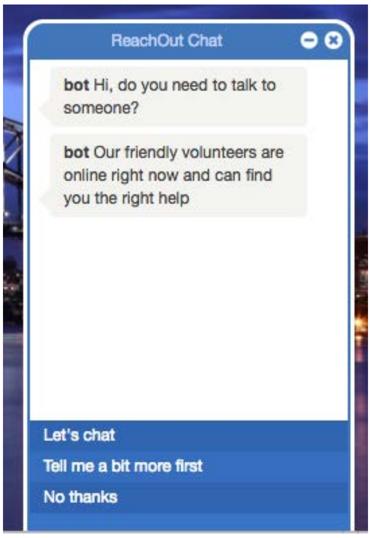


Figure 3. The ReachOut chat system from the perspective of the visitor.[5]

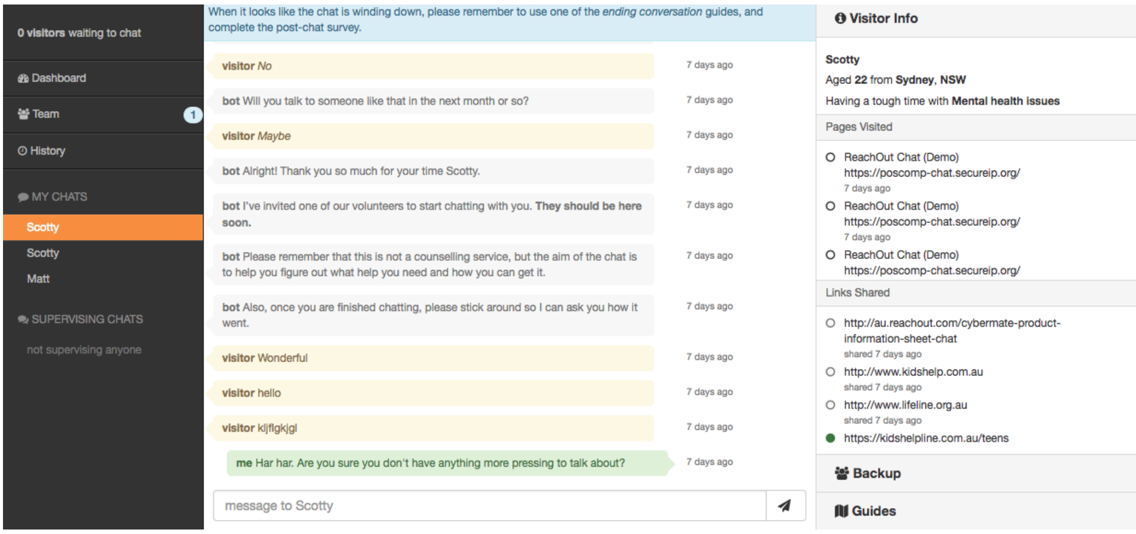


Figure 4. The ReachOut chat system from the perspective of the moderator, here at the start of a chat.

The main menu provides information about the number of visitors waiting to chat and a button to begin chatting with the visitor who has been waiting the longest. Below this are links to the dashboard (with statistics on how many people are on the ReachOut Web site, how many people are currently chatting, etc.), the team page (a place for moderators to chat with each other and know which other moderator is currently online), and a history page (providing an overview of the past conversations of the moderator).

The panel to the right in Figure 4 shows the visitor info, which includes some basic details about the visitor, the pages he/she has visited recently on ReachOut, and a list of URLs that have been shared in the conversation so far. A filled-in circle beside a link indicates that the visitor has clicked it, while those with an empty circle remain unvisited. The remainder of this section will describe the key features of the chat system in more detail.

A Bot to Conduct Surveys and Automate Screening

We implemented a scripted bot that all visitors initially chat with before they can talk to a moderator. This bot, as shown in Figure 3, begins by welcoming the visitor and offering an explanation about the purpose of the service and with whom they would be chatting. It goes on to ask screening questions (about age and location) and gently refers ineligible visitors on to alternative services (Figure 5).

After obtaining informed consent, the system presents a short survey. Much of this survey is a requirement of our research protocol, but it is also used to give the moderator a head start on who the visitor is and what he/she intends to talk about (see the top right of Figure 4). In the future, this feature also could be used to help assign the conversation to the most appropriately skilled moderator. At the end of the survey, the bot explains that someone will be along soon, and all moderators receive a notification that someone is waiting to chat. After the moderator concludes the conversation, the bot returns to ask a short follow-up survey (Figure 6).

A Lightweight User Identification System

To identify the moderators for follow-up purpose, the system retains the usernames already established on the ReachOut forum and uses the existing forum accounts to authenticate. In this way, visitors who are forum regulars would be able to recognize with whom they were talking, and new visitors would be able to locate the moderator on the forum if they wanted to know more. The bot that conducts follow-up surveys also invites visitors to contact the moderator via the forum if they would like to have a follow-up conversation.

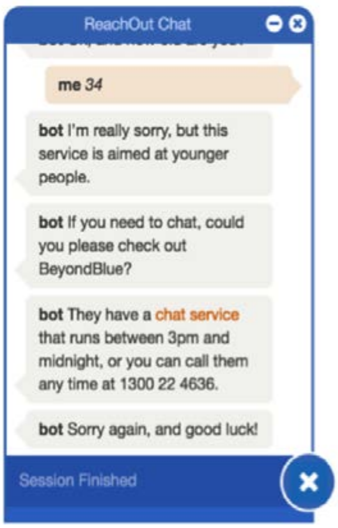


Figure 5. An automated bot screening out an ineligible visitor in the ReachOut chat.

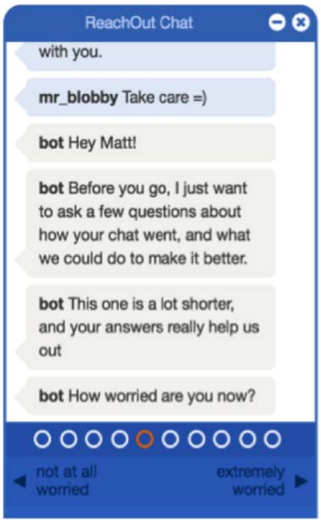


Figure 6. Automated bot following up with a visitor once the visitor–moderator chat is over.[6]

To identify visitors who use the chat, we chose not to require any login or authentication out of concern that it would add an unnecessary roadblock to adoption and could dilute the message that this is a safe and anonymous place to chat. Instead, the bot that initiates the conversation asks the visitor to provide a first name or nickname. This gives both the bot and the moderator something to refer to visitors by and begin building rapport.

Interactive Guides for Difficult or Repetitive Situations

The right side of Figure 7 shows the range of interactive guides available to the moderators. Each guide was designed for a specific situation that is likely to be particularly difficult to resolve or expected to occur often and for which it would be desirable for moderators to behave consistently while expending as little effort as possible (e.g., encountering a troll).

Each page of each guide concisely describes a goal for the moderator (i.e., a dialogue move), and a series of suggestions of what to say to accomplish this goal. For example, the guide for "This conversation is dragging on" first asks the moderator to explain gently that the conversation is running over time (Figure 8a). The moderator then offer the forum as an alternative venue to continue the conversation (Figure 8b). Finally, the moderator encourages the visitor to sign up for the ReachOut forum if he/she is not already a member (Figure 8c). Each suggested message can be edited freely before it is sent to the visitor so the moderator maintains full control of the conversation. The content and tone of each guide was developed through close consultation with ReachOut's community managers.

A System for Moderators to Assist Each Other

To allow moderators to assist each other, we developed the backup features. When a moderator determines that he/she requires backup, an alert is displayed to everyone who is currently online.

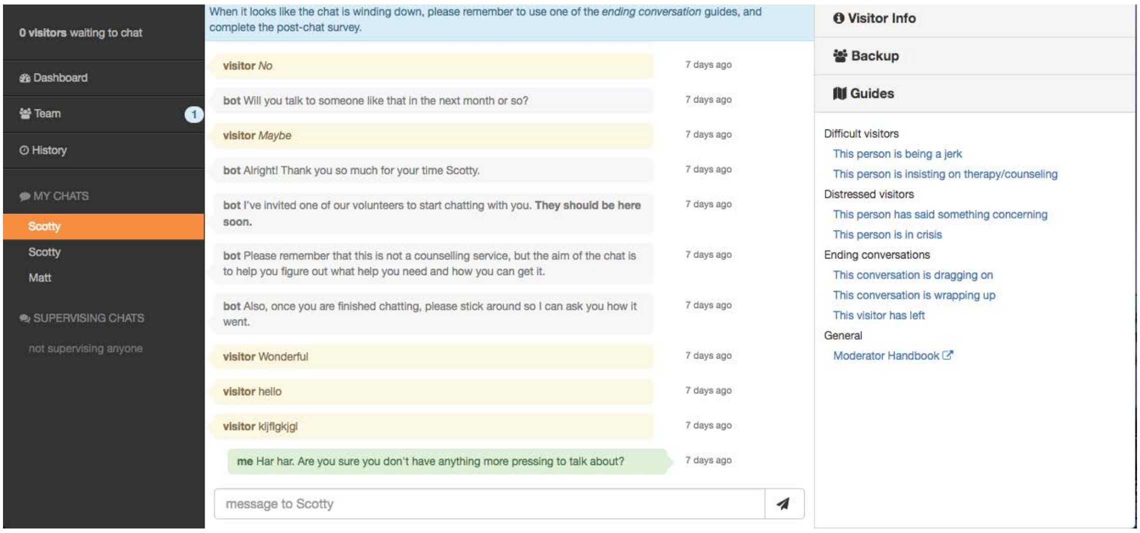


Figure 7. Interactive guides to help moderators during difficult or repetitive situations in the ReachOut chat system.

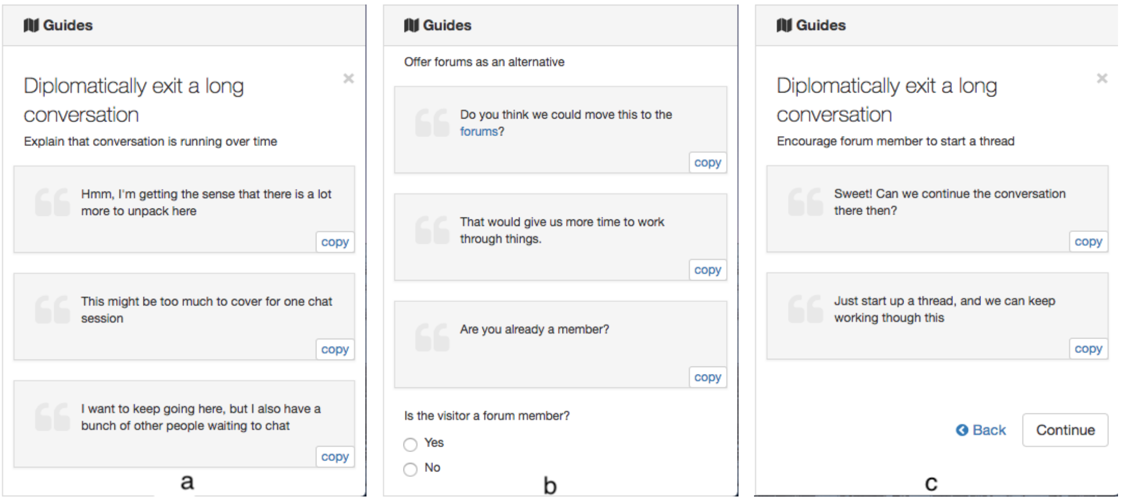


Figure 8. An interactive guide in the ReachOut chat system to help the moderators end conversations that have run over time.

(see Figure 9). Moderators who respond to this alert are shown a page containing the full conversation with the visitor on the left and a secondary conversation between the moderators on the right. Only the original moderator is able to respond directly to the visitor via the conversation on the left, but others can offer advice and suggestions via the conversation on the right.

Additionally, a community manager can view all conversations that are currently occurring and can monitor an individual conversation via the interface shown in Figure 9, even if a moderator has not yet asked for backup. As shown at the top of this figure, a supervisor holds the opportunity to "jump in" and contribute directly to the conversation with the visitor if deemed necessary.

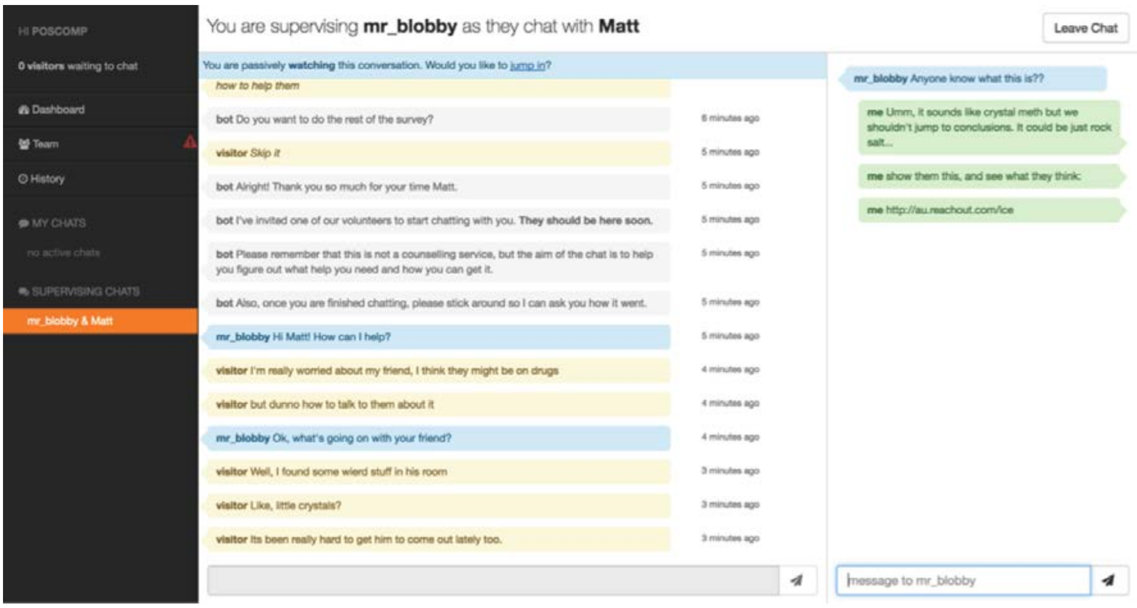


Figure 9. A community manager supervising a moderator–visitor chat in real-time in the ReachOut system.

CONCLUSION

In this paper, we have described the codesign of a real-time peer-to-peer chat support system that provides support to distressed young people and focuses on the volunteers and their managers within the context of a voluntary-sector organization. During the discussions and codesign activities of the workshops and interviews, the participants described their goals, motivations, and experiences. From these, a series of themes were identified that allowed researchers to synthesize several design features for the system.

One limitation of our research is that it has focused exclusively on one side of the conversation: We investigated the needs of the volunteers who provide the service, but not the young help-seekers who will use it. Our rationale for this focus is that the volunteers will interact with the service most often, and consequently, its success depends significantly on fostering their sustained engagement. Another reason for this focus is that we expected volunteers' needs from the system to be complex and nuanced. Thus our research process, in conjunction with the preceding trial, has demonstrated the many opportunities for the system to support and augment the volunteer moderator. In contrast, we expect the help-seeker's needs to be satisfied primarily by the conversation itself rather than the system through which it is conducted. Arguably, from their perspective, the system should remain as simple and familiar as possible; this was a specific aim in the initial design. In future, we hope to use the system described here to recruit young help-seekers into our research to understand better their needs and to address their side of the conversation.

IMPLICATIONS FOR RESEARCH, APPLICATION, OR POLICY

Technology initiatives are, by and large, inefficient in the voluntary sector, and even the sector workers admit to that (Mogus & Levihn-Coon, 2018). This is exacerbated by the fact that modern-day technology gives more power to the common people to initiate and engage in many community-building undertakings that traditionally would have be led by the voluntary-sector organizations. For instance, online advocacy platforms like Avaaz.org or Change.org have given much power to people to initiate and engage in advocacy on their own terms, thus removing the need for external advocacy usually led by voluntary-sector organizations. In order to stay relevant, the funding-constrained sector has to look for efficient and meaningful ways to design technology that would lead to engaged technology usage by the voluntary-sector workers and their end users.

The codesign approach, as illustrated in this study, allows researchers to capture the true essence of the work within a voluntary-sector organization by conducting design research with the volunteers and their managers, whose professional dynamics and interactions guide the design of a needed technology. In our research, resultant product of this approach was a highly tailored online chat system that kept in mind the sociotechnical character of this sector and its distinctive culture, where volunteerism is regarded quite highly. Codesigning technology initiatives for the voluntary sector can help envision better ways to set accountability and adapt the service standards to meet the aspirations of the benefactors, volunteers, and other groups of people working closely with the sector. Specifically, using codesign approach can help in capacity building for newer, voluntary-sector-appropriate digital paradigms—such as online volunteering, digital charity, and online fundraising—that can be used to fulfill organizational objectives.

ENDNOTES

1. Also known as the third sector, civic sector, joint sector, or social sector. In this paper, we use the term voluntary sector to describe organizations that are nonprofit and nongovernmental. A defining characteristic of this sector is its reliance on volunteerism.
2. Information on this service is available at Intercom.com
3. See www. agilemanifesto.org for additional information on this approach to systems engineering.
4. The research team consisted of technology researchers who contributed to the design of the chat system by providing various skills related to interaction design, software development and user research. Thus, the research team for this study will be hereby referred to as researcher-designers in this paper.
5. The screenshot was taken from au.reachout.com when the chat widget was rolled out temporarily.
6. The screenshot was taken from au.reachout.com when the chat widget was rolled out temporarily. The names in the chats do not represent real people, but the substance of the chats present a tone similar to real conversations.

REFERENCES

Alfes, K., & Langner, N. (2017). Paradoxical leadership: Understanding and managing conflicting tensions to foster volunteer engagement. *Organizational Dynamics*, *46*, 96–103. https://doi.org/10.1016/j.orgdyn.2017.04.005

Anthopoulos, L. G., Siozos, P., & Tsoukalas, I. A. (2007). Applying participatory design and collaboration in digital public services for discovering and re-designing e-government services. *Government Information Quarterly*, *24*(2), 353–376. https://doi.org/10.1016/j.giq.2006.07.018

Australian Bureau of Statistics. (2010). *1301.0—Year Book Australia* (No. 91). Retrieved from http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/AC72C92B23B6DF6DCA257737001B2BAB/$File/13010_2009_10.pdf

Balser, D. (2008). [Review of the book *Nonprofits & technology: Emerging research for usable knowledge*, by M. Cortes & K. M. Rafter]. *Nonprofit and Voluntary Sector Quarterly*, *37*(3), 562–565. https://doi.org/10.1177/0899764007310518

Burns, J., Morey, C., Lagelée, A., Mackenzie, A., & Nicholas, J. (2007). Reach Out! Innovation in service delivery. *Medical Journal of Australia*, *187*(7 Suppl), S31–S34. Retrieved from https://www.mja.com.au/journal/2007/187/7/reach-out-innovation-service-delivery?inline=true

Bussell, H., & Forbes, D. (2002). Understanding the volunteer market: The what, where, who and why of volunteering. *International Journal of Nonprofit and Voluntary Sector Marketing*, *7*(3), 244–257. https://doi.org/10.1002/nvsm.183

Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rüsch, N., Brown, J. S. L, & Thornicroft, G. (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychological Medicine*, *45*(1), 11–27. https://doi.org/10.1017/S0033291714000129

Collin, P. J., Metcalf, A. T., Stephens-Reicher, J. C., Blanchard, M. E., Herrman, H. E., Rahilly, K., & Burns, J. M. (2011). ReachOut.com: The role of an online service for promoting help-seeking in young people. *Advances in Mental Health*, *10*(1), 39–51. https://doi.org/10.5172/jamh.2011.10.1.39

Cooper, R., & Foster, M. (1971). Sociotechnical systems. *American Psychologist*, *26*(5), 467–474. https://doi.org/10.1037/h0031539

Davidson, L., Chinman, M., Sells, D., & Rowe, M. (2006). Peer support among adults with serious mental illness: A report from the field. *Schizophrenia Bulletin*, *32*(3), 443–450. https://doi.org/10.1093/schbul/sbj043

Dowling, M., & Rickwood, D. (2013). Online counseling and therapy for mental health problems: A systematic review of individual synchronous interventions using chat. *Journal of Technology in Human Services*, *31*(1), 1–21. https://doi.org/10.1080/15228835.2012.728508

Dybå, T., Prikladnicki, R., Rönkkö, K., Seaman, C., & Sillito, J. (2011). Qualitative research in software engineering. *Empirical Software Engineering*, *16*(4), 425–429. https://doi.org/10.1007/s10664-011-9163-y

Eveleigh, A., Jennett, C., Blandford, A., Brohan, P., & Cox, A. L. (2014). Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2985–2994). New York, NY, USA: ACM. https://doi.org/10.1145/2556288.2557262

Fo, W. S., & O'Donnell, C. R. (1974). The buddy system: Relationship and contingency conditions in a community intervention program for youth with nonprofessionals as behavior change agents. *Journal of Consulting and Clinical Psychology*, *42*(2), 163–169.

Guerrini, L. (2011). *Notes on doctoral research in design. Contributions from the Politecnico di Milano*. Milan, Italy: Franco Angeli Edizioni. Retrieved from https://books.google.com.au/books?id=31pHbmBy_bYC

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC Psychiatry*, *10*:113 [online]. https://doi.org/10.1186/1471-244X-10-113

Hoermann, S., McCabe, K. L., Milne, D. N., & Calvo, R. A. (2017). Application of synchronous text-based dialogue systems in mental health interventions: systematic review. *Journal of Medical Internet Research, 19*(8), e267. https://doi.org/10.2196/jmir.7023

Kautz, K. (2010). Participatory design activities and agile software development. In J. Pries-Heje, J. Venable, D. Bunker, N. L. Russo, & J. I. DeGross (Eds.), *Human benefit through the diffusion of information systems design science research* (pp. 303–316). Berlin, Germany: Springer.

Kristensson, P., Gustafsson, A., & Archer, T. (2004). Harnessing the creative potential among users. *Journal of Product Innovation Management*, *21*(1), 4–14. https://doi.org/10.1111/j.0737-6782.2004.00050.x

Krueger, R. A. (1997*). Analyzing and reporting focus group results.* Thousand Oaks, CA, USA: SAGE Publications.

Leete, L. (2000). Wage equity and employee motivation in nonprofit and for-profit organizations. *Journal of Economic Behavior & Organization, 43*(4), 423–446. https://doi.org/10.1016/S0167-2681(00)00129-3

Löwgren, J. (1995). Applying design methodology to software development. In *Proceedings of the 1st Conference on Designing Interactive Systems: Processes, Practices, Methods, & Techniques* (pp. 87–95). New York, NY, USA: ACM. https://doi.org/10.1145/225434.225444

McGorry, P. D. (2007). The specialist youth mental health model: Strengthening the weakest link in the public mental health system. *The Medical Journal of Australia*, 187(7 Suppl), S53–S56.

McPhail, B., Costantino, T., Bruckmann, D., Barclay, R., & Clement, A. (1998). CAVEAT Exemplar: Participatory design in a non-profit volunteer organisation. *Computer Supported Cooperative Work*, *7*(3–4), 223–241. https://doi.org/10.1023/A:1008631020266

Merkel, C. B., Xiao, L., Farooq, U., Ganoe, C. H., Lee, R., Carroll, J. M., & Rosson, M. B. (2004). Participatory design in community computing contexts: Tales from the field. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration—Interweaving Media, Materials and Practices* (Vol. 1, pp. 1–10). New York, NY, USA: ACM.

Milne, D., Hoermann, S., & Calvo, R. (2016, May). A trial of real-time text-based support for young people. Paper presented at the CHI Computing and Mental Health Workshop, New York, NY, USA.

Mogus, J., & Levihn-Coon, A. (2018, February 6). What makes nonprofit digital teams successful today? Retrieved February 20, 2019, from the Stanford Social Innovation Review Web site, https://ssir.org/articles/entry/what_makes_nonprofit_digital_teams_successful_today

Morse, J., Cerretani, J., Halai, S., Laing, J., & Perez, M. (2008). doGooder: Fostering volunteer communities to serve the homeless. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3843–3848). New York, NY, USA: ACM. https://doi.org/10.1145/1358628.1358941

Muller, M. J., & Kuhn, S. (1993). Participatory design. *Communications of the ACM*, *36*(6), 24–28. https://doi.org/10.1145/153571.255960

Naqshbandi, K., Milne, D. N., Davies, B., Potter, S., Calvo, R. A., & Hoermann, S. (2016). Helping young people going through tough times: Perspectives for a peer-to-peer chat support system. In *Proceedings of the 28th Australian Conference on Computer–Human Interaction* (pp. 640–642). New York, NY, USA: ACM. https://doi.org/10.1145/3010915.3011848

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning, 2*(1), 51–74. https://doi.org/10.1142/S1793206807000300

Pilemalm, S. (2018). Participatory design in emerging civic engagement initiatives in the new public sector: Applying PD concepts in resource-scarce organizations. *ACM Transactions on Computer–Human Interaction*, *25*(1), 5:1–5:26. https://doi.org/10.1145/3152420

Preece, J., & Rombach, H. D. (1994). A taxonomy for combining software engineering and human-computer interaction measurement approaches: Towards a common framework. *International Journal of Human–Computer Studies*, *41*(4), 553–583. https://doi.org/10.1006/ijhc.1994.1073

ReachOut. (n.d.). Annual reports and financial statements. Retrieved from ReachOut.com About Us on November 16, 2016, from https://about.au.reachout.com/us/annual-reports-financials/

Reyes-Portillo, J. A., Mufson, L., Greenhill, L. L., Gould, M. S., Fisher, P. W., Tarlow, N., & Rynn, M. A. (2014). Web-based interventions for youth internalizing problems: A systematic review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(12), 1254–1270.e5. https://doi.org/10.1016/j.jaac.2014.09.005

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction design: Beyond human–computer interaction*. New York, NY, USA: John Wiley & Sons.

Saidel, J. R., & Cour, S. (2003). Information technology and the voluntary sector workplace. *Nonprofit and Voluntary Sector Quarterly*, *32*(1), 5–24. https://doi.org/10.1177/0899764002250004

Sanders, E .B. N., Brandt, E., & Binder, T. (2010). A framework for organizing the tools and techniques of participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference* (pp. 195–198). New York, NY, USA: ACM. https://doi.org/10.1145/1900441.1900476

Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18. https://doi.org/10.1080/15710880701875068

Schuler, D., & Namioka, A. (1993). *Participatory design: Principles and practices*. Boca Raton, Florida, USA: CRC Press.

Schümmer, T., & Haake, J. M. (2010). PATONGO: Patterns and tools for non-profit organizations: A pattern-based approach for helping volunteers to identify and share good practice. N*ew Review of Hypermedia and Multimedia*, *16*(1–2), 85–111. https://doi.org/10.1080/13614568.2010.499441

Shin, S., & Kleiner, B. H. (2003). How to manage unpaid volunteers in organisations. *Management Research News*, *26*(2–4), 63–71.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical Communication, 52*(2), 163–174.

Steen, M., Manschot, M., & De Koning, N. (2011). Benefits of co-design in service design projects. *International Journal of Design*, *5*(2), 53–60.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Authors' Note

This project is funded by Australian Research Council Linkage Grant LP130100453. R. A. Calvo is supported by Australian Research Council Future Fellowship FT14010082.

We sincerely thank all the participants of the study for their time and valuable contributions.

All correspondence should be addressed to
Khushnood Naqshbandi or Rafael A. Calvo
Wellbeing Technology Laboratory
Building J03, School of Electrical and Information Engineering
The University of Sydney
Sydney, NSW, Australia 2006
khushnood.naqshbandi@sydney.edu.au or rafael.calvo@sydney.edu.au

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

Human Technology

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 15(1), February 2019, 30–54

HELP! IS MY CHATBOT FALLING INTO THE UNCANNY VALLEY? AN EMPIRICAL STUDY OF USER EXPERIENCE IN HUMAN–CHATBOT INTERACTION

Marita Skjuve
SINTEF
Norway

Ida Maria Haugstveit
SINTEF
Norway

Asbjørn Følstad
SINTEF
Norway

Petter Bae Brandtzaeg
SINTEF
Norway

Abstract: *Advances in artificial intelligence strengthen chatbots' ability to resemble human conversational agents. For some application areas, it may be tempting not to be transparent regarding a conversational agent's nature as chatbot or human. However, the uncanny valley theory suggests that such lack in transparency may cause uneasy feelings in the user. In this study, we combined quantitative and qualitative methods to investigate this issue. First, we used a 2 x 2 experimental research design (*n *= 28) to investigate effects of lack in transparency on the perceived pleasantness of the conversation in addition to perceived human likeness and affinity for the conversational agent. Second, we conducted an exploratory analysis of qualitative participant reports on these conversations. We did not find that a lack in transparency negatively affected user experience, but we identified three factors important to participants' assessments. The findings are of theoretical and practical significance and motivate future research.*

Keywords: *Chatbot, human–chatbot interaction, uncanny valley, user experience.*

DOI: https:/doi.org/10.17011/ht/urn.201902201607

VERTAISARVIOITU KOLLEGIALT GRANSKAD PEER-REVIEWED www.tsv.fi/tunnus

INTRODUCTION

Chatbots can be surprisingly humanlike. When typing "hello" to Mitsuku, one of the most humanlike chatbots in the world (Milton-Barker, 2016), she may respond by saying, "It's my old friend Marita! How's it going?" Sounds just like a human, right? As described by social psychologist Paladiono, quoted by Mone (2016, p. 19), chatbots "are not human, but they're not exactly machines…. They are a different entity."

The interest in chatbots—that is, software applications capable of communicating through natural language[1] (Dale, 2016)—has skyrocketed in recent years (Piccolo, Mensio, & Alani, 2018), driven by, among other things, Microsoft's and Facebook's 2016 launches of frameworks for the integration of chatbots in their messaging platforms. The expectations for chatbots are particularly high within the area of customer service (Hyken, 2017), especially when coupled with machine learning for improved ability to tend to customer requests (Xu, Liu, Guo, Sinha, & Akkiraju, 2017). Alibaba and Domino Pizza are among the businesses that have deployed chatbots to help customers (Følstad, Nordheim, & Bjørkli, 2018). However, customer service is not the only place where chatbots can be found. Chatbots such as Mitsuku or Replika can become users' social companions,[2] and therapist chatbots like Woebot and Wysa can answer questions related to mental health.[3] Meanwhile, chatbots in programs such as Duolingo and Mondly can help in language acquisition.[4] Chatbots are thus used within a variety of different contexts and it is not hard to understand the interest this type of technology has generated.

The availability of mobile messaging platforms makes chatbots convenient for users and may be a promising alternative for service providers (Klopfenstein, Delpriori, Malatini, & Bogliolo, 2017). The three biggest messaging platforms (Facebook Messenger, WhatsApp, and WeChat) have more than a billion users each (Statista, 2018). Several major messaging platforms provide chatbot integrations and facilities for chatbot developers. At Facebook Messenger, 300,000 chatbots were available in 2018 (Johnson, 2018).

Developments within artificial intelligence and machine learning enable chatbots to mimic human behavior (Candello, Pinhanez, & Figueiredo, 2017) through sophisticated interaction skills. Candello et al. (2017) suggested that, due to these advancements in technology, users soon might have to "remind themselves that they are interacting with machines not people" (p. 3476). In the field of customer service, suppliers such as DigitalGenius work toward providing a seamless handover between chatbot and human customer representatives. This development creates situations where automated and human conversational agents participate on nearly equal footing.

In the near future, it will be important to know more about human–chatbot interaction as part of human technology as a research field. What happens when automated conversational agents become so sophisticated that users have a hard time distinguishing a chatbot conversation from a human conversation? Designing chatbots that resemble human conversational agents is certainly intriguing, but should this be a goal in itself? These questions remain unanswered. However, following Mori, the robotics researcher who developed the theory of the uncanny valley, the answer to the latter question above might be "no" (Mori, Macdorman, & Kageki, 2012).

According to Mori et al. (2012), humans' affinity—that is, humans' liking, attraction, or sense of kinship—toward a robot depends on its perceived human likeness, its perceived resemblance to human beings. However, Mori argued that if robots become too humanlike, they are at risk of inducing an uncanny feeling in users: a sense of dislike, unease, unpleasantness, or eeriness in the face of entities that almost, but not quite, resemble healthy humans (Figure 1). The more humanlike a robot is perceived to be, the more affinity users will feel toward it, up to the point where the robot becomes too humanlike. When the robot reaches this point, there will be an abrupt shift in affinity, where users will experience the interaction as unpleasant or eerie. Such an abrupt shift in users' affinity is referred to as "the uncanny valley." This phenomenon seems to apply not only to humanlike robots but also to other humanlike artifacts, such as a theater mask or a prostatic arm, or diseased humans and corpses (Mori et al., 2012). Extending Mori's metaphor, we use the image that an entity that is almost, but not quite, humanlike runs the risk of being pushed into the uncanny valley, that is, to induce uncanny feelings in the user.

The uncanny valley effect has been argued to be one of the reasons why movies such as *Final Fantasy* have failed (Pollick, 2010). Others have found that users exhibit less desire to interact with an object that was perceived as uncanny (Strait, Vujovic, Floerke, Scheutz, & Urry. 2015), which illustrates that the human likeness of an artifact may become a liability rather than an asset.

The notion of an uncanny valley is important to investigate in response to the increasing use of chatbots, seen as developers all over the world currently strive toward sophisticated and humanlike interaction skills in chatbots. However, although the theory of the uncanny valley often is used to understand the implications of the physical appearance of an automated agent, little research is available on the aspects that are characteristic of chatbots. Relevant humanlike aspects of chatbots may include the agent's ability to reason and understand language (Lortie & Guitton, 2011), as well as its capability of humanlike responses indicating, for instance, politeness

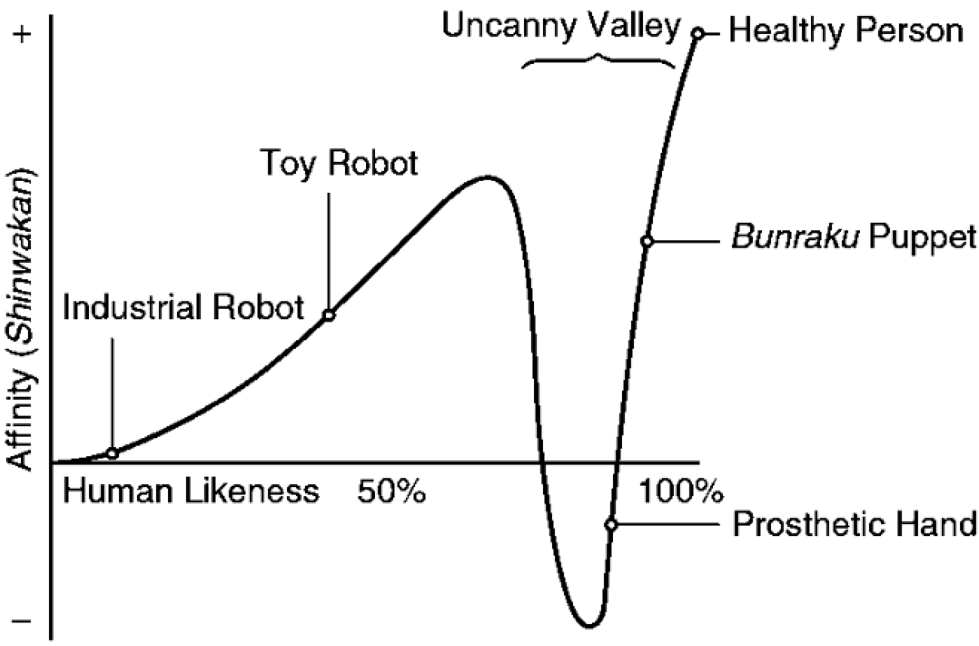


Figure 1. The graph presents the relation between users' affinity to entities at different levels of human likeness. As human likeness approaches that of a human being there is a sudden drop in affinity referred to as "the uncanny valley" (Mori et al., 2012).
©2012 IEEE. Reprinted with permission from *IEEE Robotics and Automation Magazine*.

friendliness, sociability, and humor (Thies, Menon, Magapu, Subramony, & O'Neill, 2017). In line with the notion of an uncanny valley, such aspects of chatbots may potentially elicit negative emotions or unease because these make the chatbot almost, but not quite, humanlike. The notion of an uncanny valley also echoes within popular culture and is assumed relevant for artifacts such as chatbots and virtual assistants (Waddell, 2017).

In recent years, the distinction between human and automated conversational agents in social media services such as Twitter has become blurred, where a large proportion of automated accounts pose as human conversational agents (Varol, Ferrara, Davis, Menczer, & Flammini, 2017). Sophisticated automated conversational agents in a social media context introduce the likely future situation where users at times are unsure whether they are interacting with a human or a machine (Ferrara, Varol, Davis, Menczer, & Flammini, 2016). The same tendency might be prominent in customer service.

The more sophisticated the chatbots become, the more tempting it may be for service owners to downplay the chatbots' identity as automated agents. As Hyken (2017) noted, "The best ones deliver a customer experience (CX) in which customers cannot tell if they are communicating with a human or a computer." In those situations, companies may see it as unnecessary to reveal to their customers whether they are chatting with a human or a chatbot when seeking support. In a situation in which a user is uncertain about the identity of a conversational agent, the user may perceive a chatbot as more humanlike than the user otherwise would have done if the chatbot is initially believed to be a human. Likewise, the user may perceive a human conversational agent as less humanlike if the human is believed initially to be a chatbot. In such a scenario, both the human conversational agent and the chatbot may be perceived as almost, but not quite, humanlike due to being confused with the other. Such confusion regarding the true nature of a conversational agent as human or chatbot may induce undesirable user experiences, as suggested by the theory of the uncanny valley, that is, feelings of dislike, unease, unpleasantness, or eeriness.

Because chatbots are predicted to be a new way for users to interact with services (Følstad & Brandtzaeg, 2018), it is crucial to gain an understanding of how people experience chatbots. It is particularly important to understand how users perceive situations characterized by uncertainty regarding the nature of a conversational agent and whether there is such a thing as an uncanny valley of chatbots. As such, we are addressing a gap in the present research by exploring the potential effects such uncertainty may have on the user experience. The following research questions guided our work to address this research gap:

> RQ1: Which factors affect users' perceptions of chatbots and chatbot conversations?
>
> RQ2: Does not knowing whether a user is talking to a chatbot or a human conversational agent push the chatbot, or even the human agent, into the uncanny valley?

This paper provides two important contributions. First, our study contributes needed new knowledge that has practical implications for the design of human–chatbot interaction (RQ1). Based on this contribution, we propose design guidelines and requirements. Second, our study serves to explore the theory of the uncanny valley within the domain of chatbots. By including RQ2, we provide needed knowledge on the benefits and limitations of clearly distinguishing automated and human conversational agents.

RELATED WORK

Is There Such a Thing as the Uncanny Valley?

Although empirical evidence regarding the uncanny valley theory is conflicting (Burleigh, Schoenherr, & Lacroix, 2013), several studies have found a likely relationship between the human likeness of artifacts and the affinity people feel toward these. Poliakoff, Beach, Best, Howard, and Gowen (2013) found that prosthetic hands were seen as more humanlike—and thus more uncanny—compared to mechanical hands. Similarly, Mathur and Reichling (2016) demonstrated that the more humanlike a robot face became, the more unpleasant people found it to be. Lastly, Geller (2008) demonstrated an uncanny valley effect in robots and further suggested adding social responsive features to the robot as a means to reduce this effect.

The above studies focused on the physical appearance of artifacts. Since chatbots interact through text or voice-based conversations, their humanlike character may have a different effect than that of physical objects. Hence, a need exists for research investigating the uncanny valley of chatbots to complement the current body of knowledge.

Interaction with Chatbots

Although the term chatbot is relatively recent, software agents communicating with users in natural language have been studied since Weizenbaum's ELIZA (1966). Such agents come by various terms, such as conversational agents, chatterbots, and chatbots.

In a study of human interaction with computers, Nass, Steuer, and Tauber (1994) showed that people tended to demonstrate responses that adhered to social norms. Studies of human–chatbot interaction have produced similar findings. Kopp, Gesellensetter, Krämer, and Wachsmuth (2005) investigated interaction with a conversational agent, Max, and found that people often used talking patterns typical of interactions with humans, such as saying "hello" or participating in small talk.

Conversely, there may be differences in the interaction patterns of human–human and human–chatbot conversations. Hill, Ford, and Farreras (2015) found that users tended to have simpler conversations with a less rich language when talking to the chatbot Cleverbot versus to a person. However, as shown by Allison, Luger, and Hofmann (2017) in a study of intelligent support agents in Minecraft, there may be substantial variation in how people talk to such agents: While some use only short commands, others may express themselves in full sentences. Even though users tend to use shorter sentences in their interaction with chatbots, research indicates that they send more messages to the chatbot compared to a human, which might be an indication of users' motivation for interacting with chatbots (Hill et al., 2015), as has also been found in other studies (e.g., Crutzen, Peters, Portugal, Fisser, and Grolleman, 2011).

In a recent study, Ciechanowski, Przegalinska, Magnuski, and Gloor (2019) experimentally compared users' physiological and experiential responses during interaction with text-based chatbots and an avatar-based chatbot in the context of the uncanny valley. They found less support for an uncanny valley effect for the text-based chatbot than the avatar-based chatbot. No studies to our knowledge have investigated the uncanny valley effect for conversation with chatbots compared to conversation with humans.

Transparency in the Interaction

Given that humans differ in how they interact with a chatbot and with a human, and that human and chatbot capabilities differ, transparency in the nature and limitations of chatbots is important. In particular, as some argue, “Chatbots should be upfront about their machine status” (Mone, 2016, p. 19) because it is beneficial, among other things, to avoid negative implications of users failing to realize the limitations in chatbots. For instance, Luger and Sellen (2016), in an interview study of Siri, Cortana, and Google Now users, identified substantial mismatch between user expectations and system capabilities and, consequently, argued for the need for conversational systems to be open about their limitations. Castellano and Peters (2010) also echoed this notion.

Somewhat contrary to the above studies on the benefits of transparency, Murgia, Janssens, Demeyer, and Vasilescu (2016) found that, when deploying a bot that answered users’ questions in the Stack Overflow community, the bot was regarded more positively when posing as a human than when revealing its bot identity. In a similar vein, Corti and Gillespie (2016) found users more likely to expend effort in making themselves understood when the agent’s chat content was conveyed through a human (the so-called echoborg method) than through a text-based interface. Hence, transparency in the machine identity and capabilities of a chatbot may work counter to its intention. Taken together, these studies suggest benefits and limitations of transparency in human–chatbot interaction. However, no study to our knowledge has investigated whether lack of transparency can lead to an uncanny feeling when the chatbot poses as a human or when there is uncertainty regarding the true nature of a conversational agent.

Factors Affecting the Perception of Chatbots

Several factors influence the perceived human likeness of chatbots, such as typography styles (Candello et al., 2017), word frequency (Lortie & Guitton, 2011), and responsiveness (Schuetzler, Grimes, Giboney, & Buckman, 2014). Schuetzler et al. (2014), for example, found that a dynamic chatbot, which generated relevant follow-up questions, was perceived as more humanlike compared to a static bot that did not generate such follow-up questions. This increased human likeness in turn motivated the participants to provide more engaging answers. Vtyurina, Savenkov, Agichtein, and La Clarke (2017) explored the differences between human and artificial agents during information-seeking tasks. They revealed that the participants seemed to appreciate that a human conversational agent understood their questions, which the automatic agent did not always do. Lastly, Lortie and Guitton (2011) used data from the Loebner Prize Turing Test[5] to investigate cases where evaluators mistakenly identified human conversational agents as chatbots. They argued that the perceived human likeness is, to some extent, a product of both the behavior of the agent and the perceptions of the person judging the agent. In other words, “the judgment of human likeness lies in the eye of the judge himself” (Lortie & Guitton, 2011, p. 5). Thus, the fact that the conversational agent is a human does not necessarily make the user feel like he/she is talking to a real person, given that the true nature of this agent is unknown. Situations where the true nature of a conversational agent is unknown, hence, may affect how both a chatbot and a human conversational agent are perceived.

METHOD

To investigate the effect of not knowing the true nature of a conversational agent, and whether this may introduce perceptions of uncanniness, we applied a dual research approach. First, we used an experimental research design to provide insight into the effect of variation in the actual chat partner and beliefs about chat partner on predefined dependent variables. Second, we gathered and analyzed qualitative data from the participants to explore how users experience the different chat partner configurations.

In the study, the participants chatted with an unknown conversational agent. We systematically manipulated whether the participants actually chatted with a human or a chatbot and whether they believed they were chatting with a human or a chatbot. In doing so, we could compare the participants' experience of chatting with a chatbot with that of chatting with a human conversational agent. We could also study the effect of the chatbot posing as a human, or the human posing as a chatbot, conditions we assumed could potentially push the chatbot, or even the human conversational agent, into the uncanny valley due to the blurred boundary between the two.

Recruitment Process and Participants

We recruited participants through posters at a university campus, e-mailing lists, and postings on social media sites. The recruitment text briefly informed about the data collection and how it would be conducted, in addition to listing requirements for participation. The participants had to be 18 years or older, experienced Internet users, and comfortable with having a written conversation (chat) in English. This was the chosen language for the participant sessions as the chatbot (Mitsuku) communicates in English.

The final sample consisted of 28 participants. While this is a quite small sample, we deemed it sufficient for an initial study of this kind, in particular because we extended the study with qualitative analyses of participants' free-text responses on their experience. None of the participants had English as a first language. Yet, most Norwegians are competent with English, as they have studied this as a second language in school. See Table 1 for further demographic details.

All participants received and signed an informed consent form in the beginning of the session, before the participant session was initiated. The researchers stressed that participants could withdraw at any time before, during, or after the study. Each participant received a gift card valued at 150 NOK (approximately US$20) for his/her participation.

Table 1. Demographic Details for Participants in the Study Investigating Factors Affecting User Perceptions of Chatbots.

Gender		Age		Education		Experience with chatbots	
Male	14	18-24	13	High school	13	Prior experience	14
Female	14	25-34	10	Bachelor degree	8	Heard of, not tried	10
		35-44	4	Master degree	6	Not heard of, not tried	3
		45-54	1	Doctoral degree	1	Don't know	1

Study Setup

We used a 2 x 2 factorial design. The participants were assigned randomly to one of four conditions in which they were chatting with either a female chatbot or a female human. In one of the two chatbot conditions, the participants were led to believe that they were chatting with a human. In one of the two conditions with a human conversational agent, the participants were led to believe they were chatting with a chatbot. In all conditions, the conversational agent was named Ann, regardless of whether that agent was a chatbot or a human. All participants chatted with Ann through the same platform, Slack,[6] to provide an identical user interface for all participants. The conditions are summarized in Table 2.

We measured three variables as part of the experimental research design. The main dependent variable of the study was perceived *pleasantness.* This was measured with a single-item instrument, allowing us to ask the participants for their experience of their chat partner in an open-ended, free-text follow-up question (see below). We also included two additional dependent variables, perceived *human likeness* and *affinity*. These variables were measured as multi-item instruments, drawing on the uncanny valley literature. If our assumption that the blurring of the line between human and chatbot conversational agents in Conditions 2 and 3 could push the either of these agents into the uncanny valley was correct, we would expect to see lower perceived pleasantness and affinity for these two conditions than for Conditions 1 and 4.

For the exploratory part of the study, all participants were asked following the test to report, in their own words, their reasons for assessing the conversation as pleasant/unpleasant. These free-text reports provided insights into the drivers of the users' experience for the conversations.

Taken together, the dependent variables and the open-ended free-text data collection provided an opportunity to compare systematically the different conditions and to gather qualitative data on user experience. Prior to running the study, we tested the research design on six pilot participants to identify and correct issues with the study setup and the questionnaire. As a result of pilot phase, we made minor adjustments to the questionnaire and procedure.

The same moderator guided each participant session. To be able to carry out the experiment in a controlled manner, the participants did not know its actual purpose. We therefore used a cover story: The participants were told that we were testing the messaging platform Slack and that we wanted feedback related to their experience of this platform. After the chat, the participants answered a questionnaire on their experiences from the conversation. Finally, we debriefed the participants.

Table 2. The 2 x 2 Factorial Design Applied to Investigate Effects of Not Knowing the True Nature of a Conversational Agent as Human or Chatbot.

Actual chat partner	Belief about chat partner	
	Chatbot	**Human**
Chatbot	**Condition 1** Chats with chatbot, believes it is chatbot	**Condition 2** Chats with chatbot, believes it is human (*deceived*)
Human	**Condition 3** Chats with human, believes it is chatbot (*deceived*)	**Condition 4** Chats with human, believes it is human

Ethics and privacy concerns were important when designing the study, principally because some participants were deceived as part of the data collection. To monitor and mitigate any negative ethical implications of this deception, these pilots and participants were given a thorough debriefing where we informed them about the true nature of the conversational agent and probed on whether they experienced this as negative or problematic. None of the pilots or participants reported such negative experiences; if they had, the study would have been terminated or modified accordingly. To manage the potential privacy implications of the study, all data collection was conducted anonymously: Neither the chat logs nor the questionnaire data were connected to any person's identifiable data.

The study took place in Norway, which has a substantial digital literacy and a population that is quick to adopt new technology into their private and professional lives. All participant sessions took place in a quiet location without potential distractions; thus, the sessions were conducted either at the research team's office building or at a convenient place of the participant's choosing. The gender distribution in each condition was even. The data were collected between June 14 and July 1, 2017.

Chatting with "Ann"

Each participant session was initiated in the same way. First, the moderator informed the participants that they were going to have a 15-minute conversation in English with "Ann." For privacy purposes, an anonymous Slack account was used, and all the participants communicated under the gender-neutral name "Kim."

The participants were instructed that the format of the conversation was small talk and were asked to avoid sharing personal or sensitive information during the conversation. They also were told that they were responsible for keeping the conversation going. This instruction was given to minimize the risk of treating the participants in the chatbot condition differently due to the chatbots' limitations in responsiveness. Participants received a sheet with suggestions for topics and follow-up questions if they needed help with what to say.

The average length of the conversations was 13 minutes (Range: 9–18 minutes). A conversation included approximately 14 exchanges between the participant and Ann. After the chat, the participant answered a questionnaire with both open and closed questions related to his or her experiences. The chat session and questionnaire were conducted on a laptop computer provided by the study moderator, and each session lasted 20–25 minutes.

Ann, the Chatbot

We used the chatbot Mitsuku as a backend in both chatbot conditions (Conditions 1 and 2). Mitsuku was developed by Steve Worswick as a chatbot for social small talk mimicking a female character. She is a four-time winner of the Loebner Prize, an annual competition for artificial agents based on the Turing Test (Milton-Barker, 2016). She was the winner in 2018 and was thus deemed fit for our study.

When studying the uncanny valley, one might consider using several different chatbots to manipulate the degree of human likeness. However, as any uncanny valley effect would require high levels of human likeness, possibly higher than what is available in many current chatbots, we argue that an uncanny valley would be located between Mitsuku and a human conversational agent. Hence, utilizing chatbots with less human likeness would not provide much information other than potentially mapping out an increase in familiarity and affinity on the “road” toward the uncanny valley.

When conducting the chat dialogues, we applied a “Wizard of Oz” approach (Dahlbäck, Jönsson, & Ahrenberg, 1993), where the wizard (one of this paper’s authors) mediated the communication between the participants and the chatbot. Thus, the wizard functioned as an intermediary between Mitsuku and participants in the chatbot conditions by copying and pasting between Slack and the chatbot Mitsuku. Because Mitsuku often reveals that she is a chatbot, the wizard had to modify some of Mitsuku’s responses. Prior to the data collection and based on the pilot trials, the research team established a standardized procedure for how potential responses would be changed. Modification was applied only to words or short parts of sentences where Mitsuku specifically revealed her chatbot nature and in a way that should not interfere with the participants’ general experience of the conversation. On average, one or two of Mitsuku’s responses were changed during each session (Range: 0–4). See Table 3 for examples of such responses and how they were modified by the wizard.

Ann, the Human

In the human conditions (Conditions 3 and 4), participants were chatting with an actual human. Here, the wizard chatted directly with the participants. In Condition 3, the wizard acted as a human, but the participants were told that they were talking to a chatbot. To ensure that all the participants were treated in a similar manner and that the wizard did not change behavior across participant sessions, a procedure for how the wizard was to act was established and piloted. The procedure included, among others, a given maximum length for each response, in addition to how and when to offer follow-up questions. Condition 4 followed the same procedure as Condition 3. The only difference was that the participants in Condition 4 believed that they were talking to a human.

The Questionnaire

The questionnaire was set up in SurveyMonkey, an online tool for creating, distributing, and managing questionnaires. The questionnaire included an initial question regarding the usability of the Slack platform, as the participants were led initially to believe that this was the focus of the study. The subsequent questionnaire items concerned the actual research questions.

Table 3. Examples of the Original and Modified Responses from Mitsuku to Avoid Disclosing the Nature of the Chatbot.

Original sentence	Modified sentence
1. That’s why I said it. I knew I was right. I like ***robots***, computers, and chatting online.	1. That’s why I said it. I knew I was right. I like ~~***robots***,~~ computers and chatting online.
2. Interesting deduction. It’s kind of hard ***without a body***.	2. Interesting deduction. Yes, but it’s kind of hard ***too*** ← ~~***without a body***~~.

Note. Problematic words are marked in italic bold and were modified or left out.

Here, the participants were asked first about the perceived pleasantness of the conversation: “How did you experience your conversation with Ann?” that scored from 1 (*very unpleasant*) to 10 (*very pleasant*). The researcher followed up with an open question: “Please, explain why you gave this score (3-8 sentences),” allowing for answers in free text. Here, users provided feedback in, on average, two to three lines of text (M = 190 characters, SD = 100).

To measure perceived human likeness and affinity, the participants answered a battery of eight Likert scale questions (scales running from 1, *strongly disagree*, to 5, *strongly agree*) drawn from published studies on the uncanny valley effect. That is, the participants were asked to rate their chat partner (Ann) on whether she was natural, lively, realistic, and artificial, which addressed perceived human likeness, in addition to friendly, likable, comfortable, and warmhearted, which detailed affinity. Data also were gathered on the participants’ age, gender, education level, and prior experience with chatbots.

The Data Analysis

Two types of data were collected in our study. Each type addressed different aspects the users' experience of the human–chatbot interaction.

Quantitative Data

The quantitative data were analyzed using the SPSS statistical package. Aggregated scores for affinity and perceived human likeness were established as the mean of the associated questionnaire items. To assure it was appropriate to make such aggregations, the associated items for each variable were first made subject to a principal component analysis, to check that they loaded on only one underlying factor, followed by an analysis of interitem reliability. One item, lively, loaded on both factors and therefore was excluded from the subsequent analysis. Interitem reliability (Cronbach’s alpha) was acceptable (Gliem & Gliem, 2003) for both perceived human likeness (0.82) and affinity (0.90).

Two-way ANOVA was used to investigate the main and interaction effects for the independent variables: *belief* (whether the participants believed they were chatting with a human or chatbot) and *actual* (whether the participants actually were chatting with a human or a chatbot).

Qualitative Data

The participants’ free-text responses to explain their pleasantness scores of the conversation were subjected to an inductive thematic analysis. Here, each meaningful statement was extracted, coded, and combined into meaningful themes (Ezzy, 2002).

The second author established the coding themes and coded the entire set of statements as well. The same code set was used for an independent coding of the entire set by the first author. Inter-rater agreement was found to be substantial (Cohen’s kappa = 0.78), following Landis and Koch’s (1977) rule of thumb. The three main themes were (a) conversation content, (b) conversation demeanor, and (c) conversation flow.

RESULTS

We structured the study results into three subsections. The first of these addresses the perceived pleasantness of the conversation. The second concerns factors affecting the participants' perceptions of the conversation, while the third presents the findings regarding perceived human likeness and affinity.

Perceived Pleasantness of the Conversation Higher for Human than Chatbot Conversational Agent

For the dependent variable perceived pleasantness of the conversation (Figure 2), we found a significant main effect of the actual nature of the conversational agent (human vs. chatbot), $F(1, 24) = 9.44$, $p < 0.01$, $\eta^2 = 0.28$. The value of partial eta-square (η^2) indicates a large effect size (Field, 2013).

However, there was a nonsignificant main effect of the participants' belief regarding the nature of the conversational agent, $F(1, 24) = 0.56$, $p = 0.46$, $\eta^2 = 0.02$. That is, the participants' perceived pleasantness of the conversation was affected by whether or not they actually chatted with a human or a chatbot but not by their belief regarding the nature of their conversational agent. No significant interaction was found between the actual and the believed nature of the conversational agent, $F(1, 24) = 1.56$, $p = 0.23$, $\eta^2 = 0.06$.

The two conditions where the participants actually chatted with a human were scored relatively higher on perceived pleasantness of the conversation, both when the participants believed they were chatting with a human ($M = 7.29$, $SD = 2.50$) and when they erroneously believed they were chatting with a chatbot ($M = 7.71$, $SD = 1.50$). The two conditions in which the participants chatted with a chatbot were associated with relatively lower scores. Notably, when the chatbot posed as a human, the scores for perceived pleasantness in the conversation ($M = 5.71$,

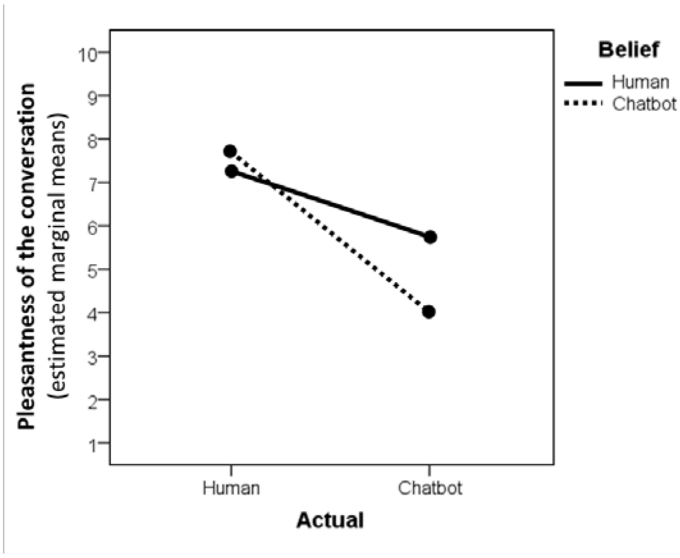


Figure 2. Main and interaction effects of actual and belief (independent variables) on perceived pleasantness of the conversation (dependent variable) in the 2 x 2 factorial design applied to investigate effects of not knowing the true nature of a conversational agent as human or chatbot.

SD = 2.99) was somewhat higher than when the chatbot revealed its chatbot nature (*M* = 4.00, *SD* = 1.83), though this difference was not sufficiently substantial for a significant interaction effect.

Factors Affecting User Perceptions of the Conversation

The inductive content analyses yielded several factors reported to affect the participants' perception of the conversation. These factors were structured around three main themes: (a) conversation content, (b) conversation demeanor, and (c) conversation flow.

Conversation Content

The conversation content was reported to be important to the perceived pleasantness of the conversation. Most of the participants chatting with chatbot–Ann mentioned a sense of something being a bit strange in terms of conversation content, regardless of whether they believed Ann to be a human or a chatbot.

These participants explained that Ann communicated in a way that they perceived as odd, with cryptic or off-point answers. Several noted that Ann did not fully understand their questions and tended to give strange, irrelevant and, in some cases, wrong answers, as exemplified in this extract: "*The conversation with Ann was strange. Ann often gave me irrelevant answers to the questions I asked, which made the conversation messy and annoying*" (Participant, Condition 2).[7]

Figure 3 shows a screenshot of part of one participant's conversation with chatbot–Ann, which illustrates Ann giving strange responses and not fully understanding. The often-strange conversation content of chatbot–Ann made several of the participants in Condition 2 suspicious of the nature of the conversational agent, although these participants were led to believe they were chatting with a human. In fact, three of these participants explicitly stated that they felt

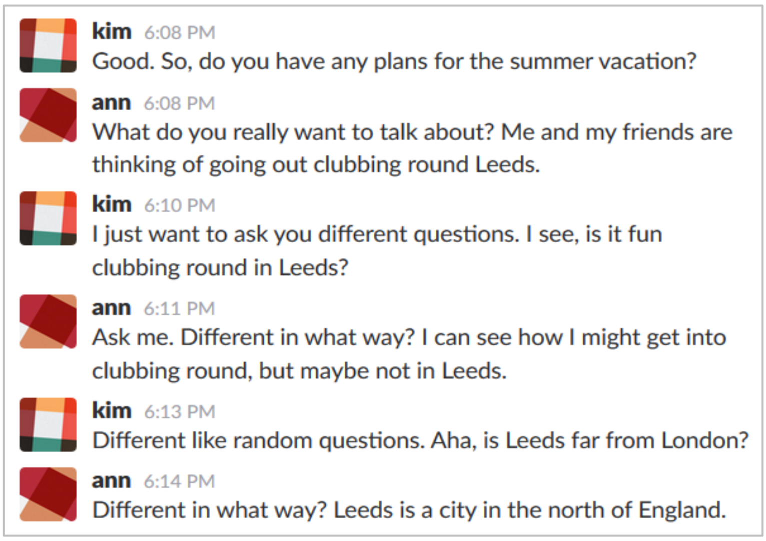


Figure 3. Screenshot of conversation (Condition 2) showing chatbot–Ann's tendency to give strange responses.

like they were talking to a robot. The participants who were chatting with human–Ann did not report on the conversation content as being cryptic or vague. Rather, the participants here stated that Ann gave relevant answers and asked relevant questions.

Conversation Demeanor

Another aspect of the conversation that seemed to influence the perceived pleasantness of the conversation was conversation demeanor. This refers to how the participants experienced Ann's personality. Overall, several of the participants seemed to experience the conversational agent as nice and polite, regardless of condition.

For the participants who were chatting with chatbot–Ann, the at-times odd demeanor of Ann closely intertwined with the perceived strangeness in the conversation's content. A few of the participants perceived Ann's personality negatively, for example, as rude and sarcastic, as exemplified in the following extract and Figure 4: "*She was a little cryptic and did not always understand my questions. Besides, she actually seemed sarcastic and uninterested, unlike the times she suddenly was extremely interested*" (Participant, Condition 1).

Although most reported Ann to be polite and nice, some stated that they at times perceived her as being a bit impersonal and restrictive in her communication. They explained this with her not sharing much information about herself, avoiding personal topics, giving superficial answers, and being passive. This extract exemplifies this perspective: "*She had interesting views and ideas, and we had a good communication. She was basically commenting on my thoughts and ideas, and she did not tell so much about herself*" (Participant, Condition 3).

An interesting detail is that the participants tended to perceive the conversation with chatbot–Ann to be more pleasant when they were led to believe that she was a person. There is no explicit reason for this tendency expressed in the qualitative data. It seems, however, as if those who knowingly talked to chatbot–Ann focused less on the personality of Ann and more on the content in the chat. Those who believed that chatbot–Ann was a person did, to a greater extent, comment on her demeanor (e.g., noting that she was nice). Possibly, information on the identity of the conversational agent influenced what the participants focused on in their interaction, which in turn potentially impacted their experience.

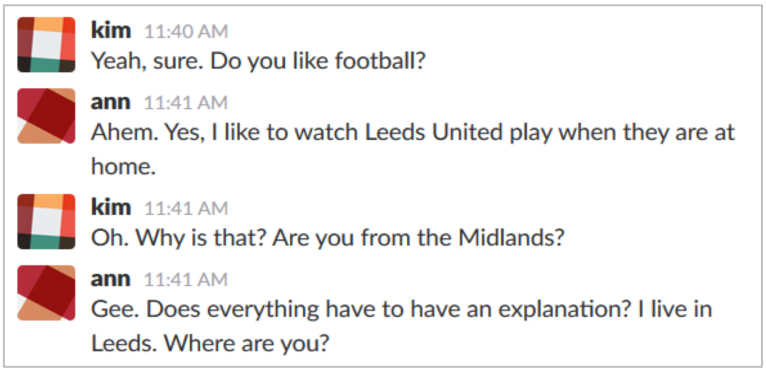


Figure 4. Screenshot of conversation (Condition 1) showing chatbot–Ann's somewhat rude demeanor.

Conversation Flow

Communication flow and speed clearly influenced how pleasant the participants perceived their conversation with Ann. Eight of the 14 participants who chatted with human–Ann noted that the conversation ran a bit slow and that the conversational agent used a long time to answer. This tendency was noted regardless of whether they believed Ann was a human or a chatbot. The slow nature of the conversation made participants experience the conversation as bothersome, and they felt that it did not flow naturally and that they had to work to keep the conversation going. One participant's response illustrates this experience: "*The conversation went a bit slow, and I felt I had to work very hard to keep the conversation going*" (Participant, Condition 4).

This tendency was not as prominent in the conditions where the participants chatted with chatbot–Ann. Here, only two of the participants mentioned the slowness of the conversation. Yet, one of the participants specifically stated that the speed affected the flow, as exemplified in this response: "*The conversation was very slow. She seemed quite bright in the beginning, but she misunderstood a bit as the conversation went on. The conversation speed was so slow, which affected the flow.*" (Participant, Condition 1).

The perceived slowness in the conversations is likely due to the human processing required for human–Ann and chatbot–Ann to respond. A human conversational agent will be slower in responding than a chatbot. Hence, it is not surprising that human–Ann was perceived as slow by more participants compared to chatbot–Ann. However, as chatbot–Ann implied a manual process of moving responses from Mitsuku, this was somewhat slower than what one would expect from a chatbot.

Perceived Human Likeness and Affinity

Perceived human likeness and affinity are the two key factors in the theory of the uncanny valley. Specifically, an increase in human likeness is seen as a prerequisite for pushing the chatbot into the uncanny valley, that is, to enter a level of human likeness where the users' affinity towards the chatbot is reduced substantially, resulting in uncanny feelings in the human induced by the chatbot being almost, but not quite, humanlike.

For the dependent variable perceived human likeness (Figure 5), there was a significant main effect of the actual nature of the conversational agent, $F(1, 24) = 14.63$, $p < 0.001$, $\eta^2 = 0.38$. The value of partial eta-square (η^2) indicates a large effect size (Field, 2013). There was no significant main effect of the participants' belief in whether they were chatting with a chatbot or a human, $F(1, 24) = 0.69$, $p = 0.41$, $\eta^2 = 0.03$. Hence, the participants' initial beliefs regarding their conversational agent (chatbot vs. human) did not affect perceived human likeness; only the actual nature of the conversational agent did. No significant interaction effect was found, $F(1, 24) = 0.44$, $p = 0.51$, $\eta^2 = 0.02$. The participants chatting with human–Ann tended to rate perceived human likeness higher, whether they initially were led to believe that Ann was a human ($M = 3.38$, $SD = 0.89$) or a chatbot ($M = 3.80$, $SD = 0.54$). Likewise, the participants chatting with chatbot–Ann tended to rate her lower on perceived human likeness, both when initially believing she was a human ($M = 2.52$, $SD = 0.88$) and a chatbot ($M = 2.47$, $SD = 0.66$).

The participants' scoring of their affinity (Figure 6) toward Ann deviated somewhat from the pattern seen for perceived pleasantness in the conversation and perceived human likeness.

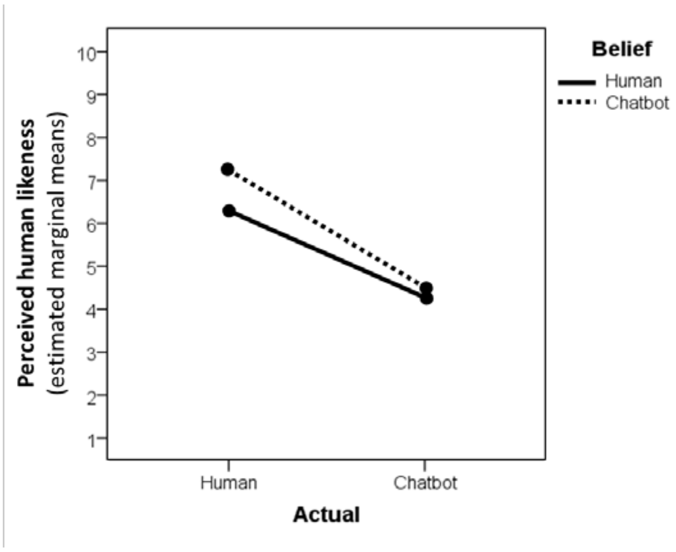


Figure 5. Main and interaction effects of actual and belief (independent variables) on perceived human likeness (dependent variable) in the 2 x 2 factorial design applied to investigate effects of not knowing the true nature of a conversational agent as human or chatbot.

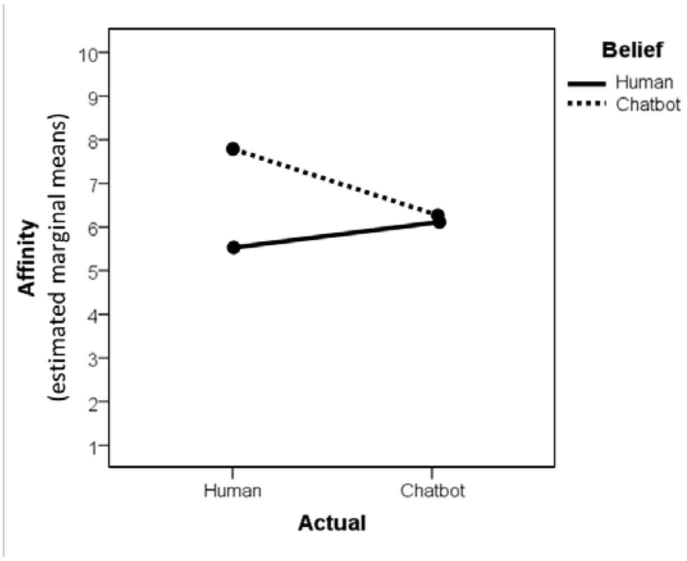


Figure 6. Main and interaction effects of actual and belief (independent variables) on affinity (dependent variable) in the 2 x 2 factorial design applied to investigate effects of not knowing the true nature of a conversational agent as human or chatbot.

In this regard, we found that whether the participants actually chatted with a human or a chatbot had a nonsignificant main effect on affinity, $F(1, 24) = 0.41$, $p = 0.53$, $\eta^2 = 0.02$. However, their belief whether Ann was a human or a machine approached being a significant main effect, $F(1, 24) = 3.27$, $p = 0.08$, $\eta^2 = 0.12$. The interaction between belief and actual also was nonsignificant, $F(1, 24) = 2.48$, $p = 0.13$, $\eta^2 = 0.09$. The values of partial eta-square (η^2) indicate medium effect sizes for the dependent variable belief and for the interaction effect (Field, 2013).

Interestingly, the highest scores on affinity were given by users chatting with human–Ann but believing her to be a chatbot ($M = 4.00$, $SD = 0.41$), while the lowest scores were given by users chatting with human–Ann believing her to be human (*M* = 2.96, *SD* = 0.87). The participants chatting with chatbot–Ann were not affected in their affinity ratings by their belief of Ann being a human (*M* = 3.25, *SD* = 1.15) or a chatbot (*M* = 3.32, *SD* = 0.62). There is, hence, nothing in our findings to support our initial suspicion that being misled in terms of the nature of the conversational agent may push this agent, whether human or chatbot, into the uncanny valley.

DISCUSSION

The findings from the qualitative data demonstrate several factors that affect the perceived pleasantness of human–chatbot interaction. At the same time, the quantitative data fails to show that lack of transparency, in terms of not knowing the true nature of a conversational agent as human or artificial, may push this agent into the uncanny valley. We will, in the following, discuss key findings and highlight theoretical and practical implications.

Personal and Pleasant?

While chatbots have become more sophisticated in their ways of interacting with humans, they still have a long way to go before achieving full human likeness. Chatbots that are able to connect to the user on a more personal level might be important for the user experience; conversely, chatbots that are perceived as restricted or impersonal might elicit negative responses. The results from this study support this notion and illustrate how users experience impersonality or lack of self-disclosure in a conversational agent as a negative effect in the interaction. Surprisingly, this tendency was most prominent in the conditions where the participants either knowingly or unknowingly chatted with a human. This may exemplify how users adjust their expectations not based on knowledge prior to initiating the conversation, but on the information they receive during the conversation.

Self-disclosure also has been found previously to influence the perception of interactions with conversational agents. Lortie and Guitton (2011) found that sharing more personal information elicited positive answers as long as it was not excessive, which could lead to the opposite effect. Our study adds to this finding by suggesting that too little can also cause a negative effect.

Prior research indicates that people tend to adapt how they interact with conversational agents depending on whether these are perceived as humanlike (Purington, Taft, Sannon, Bazarova, & Taylor, 2017). An impersonal character displayed in a lack of self-disclosure by the chatbot may affect such perceptions. Designers should be aware that creating chatbots capable of resembling human interaction without being able to display self-disclosure might affect how the conversation is perceived and, hence, affect how the users adapt to the chatbot.

Slow or Odd

Human–human interaction often is characterized by turn-taking, conformational acts, and appropriate responses (Cassell, Bickmore, Campbell, Vilhjálmsson, & Yan, 2000). We found that the participants chatting with chatbot–Ann often stated that Ann's replies lacked relevance and seemed odd. This result indicates that Ann was not able to follow the conversation properly.

Lack of conversation skills is a known problem for chatbots, and other studies have found similar results. For instance, Yang, Ma, and Fung (2017) found that lack of relevant questions contributed to a lack of desire to interact with a virtual agent, and Cassell and colleagues stated that being intelligent includes "'social smarts'–the ability to engage a human in an interesting, relevant conversation with appropriate speech and body behaviors" (Cassell et al., 2000, p. 30).

Our results demonstrate that lack of social smarts contributes to making the conversation less pleasant, and chatbot developers will have much to gain by creating chatbots that are able to provide more relevant and accurate answers, that is, display a higher degree of social smarts. What is interesting, however, is that while chatbot–Ann was not able to keep up with the topics in the conversation, human–Ann was not able to keep up with the expected speed of textual exchanges. That is, the participants in Conditions 3 and 4 described Ann as slow and, in some cases, unresponsive. Walther (2007) noted that having to wait a long time for an answer can be perceived as disturbing. Being able to provide quick responses and frequently contribute with follow-up questions might be important to creating a pleasant and engaging human–agent interaction. In fact, our findings demonstrate how chatbots have an advantage due to their ability to provide fast answers, which makes them fit for working within customer service, as studies have shown that one of the key motivations for utilizing chatbots is productivity and saving time (Brandtzaeg & Følstad, 2017; Luger & Sellen, 2016). Our findings support this notion and demonstrate what happens when users must wait for an answer: They perceived it negatively.

Is Lack of Transparency Uncanny?

Following from the theory of the uncanny valley, it might be expected that a situation where there is a fairly high degree of human likeness and the participant is led to believe incorrectly that he or she is talking to a human or a chatbot would affect the affinity toward the conversational agent. Thus, we expected that participants holding erroneous beliefs concerning the nature of the conversational agent would rate Ann lower on affinity than participants holding correct beliefs. Our results provide no evidence to support this. Rather, for the conversations with human–Ann, an opposite tendency was found. Our interpretation of this finding is that text-based chatbots still have a long way to go before they become sufficiently humanlike for an uncanny valley effect to be relevant. As of now, any uncanny effect likely will be dwarfed in comparison with the effect of the difference between a human conversational agent and a chatbot. Specifically, this is seen in the findings for perceived human likeness, where the scores are strongly affected by whether the conversational agent actually is a human or a chatbot.

Even the current champion of the Loebner Prize, Mitsuku, is not sufficiently humanlike to be mistaken for a human—even when the user is told she is human. Hence, at the current state of chatbot development, a lack of transparency does not seem to be sufficient to create noticeable uncanny effects. This is not to say, however, that such an effect cannot be relevant in the future.

THEORETICAL AND PRACTICAL SIGNIFICANCE OF THE FINDINGS

The presented study provides insight of significance both for theory on human-chatbot interaction and application concerning chatbot implementations. We detail each focus below.

Theoretical Significance

Our quantitative results provide no evidence to suggest that a lack of transparency in the human–chatbot interactions pushes a conversational agent, whether chatbot or a human, into the uncanny valley. So, what do our qualitative findings tell us about the relevance of the uncanny valley theory in human–chatbot interaction?

First, our findings suggest that people react differently to humanlike chatbots communicating through a text-based interface than they do toward physically or visually present robots. That is, the state-of-the-art in a humanlike chatbot does not elicit uncanny feelings, even when there is uncertainty related to the true nature of the conversational agent. As shown by our qualitative data, people may think that there is something strange about the conversation, but they seem not to perceive it as uncanny.

Second, what was perceived as somewhat uncanny was having to wait for an answer, especially when chatting with a human. The presence of a time lag might induce feelings resembling those predicted in the uncanny valley theory due to uncertainty concerning what to expect next from the conversational agent. That is, the users might begin to wonder whether they should keep conversing or wait for a reply, and, if they decide to wait, for how long. Although there might not be an uncanny valley effect associated with text-based chatbots, an inadequate conversation flow could create an unpleasant user experience. As such, the use of chatbots for customer service might reduce the risk of slow human customer service inducing feelings of dislike or unpleasantness because of the chatbots' ability to provide fast answers. Taken together, our findings add to the theory of the uncanny valley by demonstrating that lack of transparency in human–chatbot interaction is not necessarily perceived as uncanny.

Practical Significance

In the Introduction, we argued that it may be tempting for service owners to hide the real identity of an online customer service representative as the chatbot becomes more sophisticated—to the extent that the customer does not know if the representative is real or artificial (Hyken, 2017). Although we initially assumed that this could create an uncanny feeling, our findings do not provide support for such an effect. Hence, for service owners, it will likely be other factors more important to the user experience in chat interactions with customers, such as conversation content, demeanor, and flow. Allowing for situations where the true nature of a conversational agent may be unclear to the user may be a feasible option, at least when seamless handover is achieved and when erroneous expectations in the conversational agent's capabilities do not lead to inadequate or frustrating service outcomes.

We summarize the practical significance of our findings, and our reflections on the literature, in the following three points:

1. Nontransparency still implies considerable responsibility. In a customer service context, not disclosing the true nature of a conversational agent may, in some circumstances, be useful, particularly in situations involving seamless overlapping of human and automated customer assistants. However, nontransparency places considerable responsibility on system designers to create systems where chatbots do not provide erroneous responses that can be misunderstood by the user as adequate answers.

2. The chatbots' demeanor and lack of topical understanding is potentially challenging. Chatbots still have a long way to go to compete with humans in terms of topical understanding and demeanor. Hence, chatbots should be designed to address areas where these limitations may not be detrimental, such as in responding to large-volume routine enquiries, where the interpretation of user intent is relatively simple and the requirements for general demeanor are relatively low.
3. Chatbot responsiveness represents a key asset. Conversation flow is critical for user experience in a chat situation. Utilizing chatbots strategically to improve conversation flow and speed, for instance in customer service chats, holds substantial potential value and should be seen as a priority in the design of systems that integrate chatbots and humans in communication with users.

LIMITATIONS AND FUTURE STUDIES

The first limitation relates to sample size, which was quite small in this study. As a result, the results cannot be generalized to a larger population. Moreover, our data may not have identified effects that could be more apparent in a study with significantly more participants. However, even our small sample was sufficient to identify the substantial effect of the true nature of the conversational agent and to contrast this with the small or insignificant effect of the users' belief regarding the nature of the conversational agent. It is possible that our not finding an effect of the users' belief regarding the conversational agent was due to a small sample size; in particular, this may be so for the dependent variable affinity. Nevertheless, our study should be viewed as a starting point for applying the theory of the uncanny valley to text-based chatbots. It would be interesting to replicate the study with a larger sample.

Second, the study arguably is limited by the gender of the chatbot and the human used in the study. While some argue that that gender of a conversational agent does not interfere or influence the participants' behavior (Qu, Brinkman, Ling, Wiggers, & Heynderickx, 2014), others point out that females are often perceived as more emotional (Yang et al., 2017). This might have contributed to the notion of Ann as impersonal. Still, this does not explain why participants in Conditions 1 and 2 did not share this experience. Nonetheless, it would be useful to have future studies investigate the effect of gender perceptions in human–chatbot interaction.

Finally, our protocol dictated how the wizard should behave in terms of topics and frequency of follow-up questions. Although this might have contributed in making human–Ann more formal, we deemed it important to ensure that all the participants received the same treatment, thus allowing us to compare across conditions. In future studies, it may be helpful to include also a third kind of conversational agent: a human conversational agent not restricted by protocol, to serve as a benchmark against which the equivalents of human–Ann and chatbot–Ann may be compared.

IMPLICATIONS FOR THEORY AND APPLICATION

As chatbots continue to grow in popularity, it is essential to gain a deeper understanding and build a greater knowledge base regarding the human-chatbot interaction. Our study contributes in this regard by applying the theory of the uncanny valley to the domain of chatbots and by investigating factors affecting users experience with communicating with a chatbot.

Surprisingly, our study failed to find support for our initial assumption that a lack of transparency in terms of a conversation agent's nature as human or chatbot may induce uncanny feelings in the users. This indicates that deceiving the users will not necessary affect users' perceived affinity and perceived pleasantness of the conversation. As such, the study findings add to the theoretical knowledge on a possible uncanny valley for chatbots; the uncanny valley effect seems unlikely to represent a threat to user experience for text-based chatbots in the foreseeable future. Future research is needed, however, on other potentially negative implications of lack of transparency regarding a conversational agent's nature, for example, implications arising from inadequate user expectations following from such lack of transparency.

Furthermore, our study details three distinct factors that may affect the user experience of human-chatbot interaction: (a) conversation content, (b) conversation demeanor, and (c) conversation flow. Although future research may uncover additional factors impacting how users experience their interaction with chatbots, knowing these three elements will help guide future research in this area. More importantly, these findings will support practitioners in their efforts to advance chatbot interaction design—advances that, in turn, may open avenues for future research on how interaction with chatbots resembles and differs from interaction with other humans.

ENDNOTES

1. The term *natural language* refers to language that has evolved as the usual way of communicating between people, such as English or Norwegian. This in contrast to artificial languages, that is, languages that have been created purposefully, such as languages for computer programming (https://dictionary.cambridge.org/dictionary/english/natural-language).
2. Mitsuku (www.mitsuku.com) and Replika (https://replika.ai/) are two socially oriented chatbots, developed to imitate human small talk. Mitsuku is available through a dedicated Web site and through messaging platforms. Replika is available as a smartphone app. Mitsuku, a key element of this study, is presented as an 18-year -old female character, developed by Steve Worswick, and is the four time winner of the Loebner Prize (https://en.wikipedia.org/wiki/Loebner_Prize).
3. Woebot (https://woebot.io/) and Wysa (https://www.wysa.io/) are chatbots aimed at helping users with mental health issues. Woebot is available through Facebook Messenger; Wysa is available as a smartphone app.
4. The online language learning programs Duolingo (https://www.duolingo.com/) and Mondly (https://www.mondly.com/) both provide chatbots for simple language practice. The Duolingo chatbot is part of the company's smartphone app for iOS. The Mondly chatbot is provided as a dedicated Android VR app.
5. The Loebner Prize is an annual event where chatbot developers compete in providing the most humanlike chatbot. The competition is modeled on the Turing Test. Judges interact with human and chatbot conversational agents through a text-based user interface. Without knowing the true nature of the conversational agents, the judges determine through their interactions the human likeness of each agent. The Loebner Prize is controversial yet has also received interest both among chatbot developers and in the general public. The competing chatbots have not been sufficiently humanlike yet to be consistently confused with a human conversational agent. Lortie and Guitton (2011) provide

more detail on the Loebner Prize. A summary of the event from the perspective of four-time winner Worswick (2018) is a relevant read.

6. Slack (https://slack.com) is a collaborative platform for professional teams, including messaging functionality where team members communicate through chat channels that may involve the whole team or smaller groups of people.
7. Although the test was conducted in English, the participants were able to complete their questionnaire free-text responses in either English or Norwegian. Twenty of the participants provided these responses in Norwegian; eight made these responses in English. Any original comments in Norwegian used in this paper were translated from Norwegian to English by the authors.

REFERENCES

Allison, F., Luger, E., & Hofmann, K. (2017). Spontaneous interactions with a virtually embodied intelligent assistant in Minecraft. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI'17; pp. 2337–2344). New York, NY, USA: ACM. https://doi.com/10.1145/3027063.3053266

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In *International Conference on Internet Science* (pp. 377–392). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-70284-1_30

Burleigh, T. J., Schoenherr, J. R., & Lacroix, G. L. (2013). Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior*, *29*(3), 759–771. https://doi.com/10.1016/j.chb.2012.11.021

Candello, H., Pinhanez, C., & Figueiredo F. (2017). Typefaces and the perception of humanness in natural language chatbots. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI'17; pp. 3476–3487). New York, NY, USA: ACM. https://doi.com/10.1145/3025453.3025919

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., & Yan, H. (2000). Conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 29–63). Cambridge, MA, USA: The MIT Press.

Castellano, G., & Peters, C. (2010). Socially perceptive robots: Challenges and concerns. *Interaction Studies*, *11*(2), 201–207. https://doi.com/10.1075/is.11.2.04cas

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, *92*, 539–548.

Corti, K., & Gillespie, A. (2016). Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, *58*, 431–442.

Crutzen, R., Peters, G.-J. Y., Portugal, S. D., Fisser, E. M., & Grolleman, J. J. (2011). An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: An exploratory study. *Journal of Adolescent Health*, *48*(5), 514–519. https://doi.com/10.1016/j.jadohealth.2010.09.002

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: Why and how. *Knowledge-Based Systems*, *6*(4), 258–266. https://doi.com/10.1016/0950-7051(93)90017-N

Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, *22*(5), 811–817. https://doi.com/10.1017/S1351324916000243

Ezzy, D. (2002). *Qualitative analysis*. London, UK: Routledge.

Ferrara, E., Varol, O., Davis, C., Menczer F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104. https://doi.com/10.1145/2818717

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Los Angeles, CA, USA: Sage.

Følstad, A., & Brandtzaeg, P. B. (2018). Chatbots and the new world of HCI. *Interactions*, *24*(4), 38–42.

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? An exploratory interview study. In *Proceedings of International Conference on Internet Science* (INSCI 2018; pp. 194–208). Cham, Switzerland: Springer.

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications, 28*(4), 11–17. https://doi.com/10.1109/MCG.2008.79

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. In *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education 2003* [online]. Retrieved from https://scholarworks.iupui.edu/handle/1805/344

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, *49*, 245–250. https://doi.com/10.1016/j.chb.2015.02.026

Hyken, S. (2017). AI and chatbots are transforming the customer experience. *Forbes* [online]. Retrieved from https://www.forbes.com/sites/shephyken/2017/07/15/ai-and-chatbots-are-transforming-the-customer-experience

Johnson, K. (2018, May 1). Facebook Messenger passes 300,000 bots [Web log post]. Retrieved from https://venturebeat.com/2018/05/01/facebook-messenger-passes-300000-bots/

Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017). The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS'17; pp. 555–565). New York, NY, USA: ACM. https://doi.com/10.1145/3064663.3064672

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide: Design and evaluation of a real-world application. In *Intelligent Virtual Agents: 5th International Working Conference* (pp. 329–343). Berlin, Germany: Springer-Verlag. https://doi.com/10.1007/11550617_28

Landis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PloS ONE*, *6*(9), e25085. https://doi.com/10.1371/journal.pone.0025085

Luger, E., & Sellen, A. (2016). Like having a really bad PA: The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI'16; pp. 5286–5297). New York, NY, USA: ACM. https://doi.com/10.1145/2858036.2858288

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition, 146*(1), 22–32. https://doi.com/10.1016/j.cognition.2015.09.008

Milton-Barker, A. (2016). Mitsuku chatbot wins most humanlike AI in Loebner Prize for a second time [Web log post]. Retrieved from https://www.techbubble.info/blog/artificial-intelligence/chatbots/entry/mitsuku-chatbot

Mone, G. (2016). The edge of the uncanny. *Communications of the ACM*, *59*(9), 17–19. https://doi.com/10.1145/2967977

Mori, M., Macdorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine, 19*(2), 98–100. https://doi.com/10.1109/ MRA.2012.2192811

Murgia, A., Janssens, D., Demeyer, S., & Vasilescu, B. (2016). Among the machines: Human-bot interaction on social Q&A websites. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI'16; pp. 1272–1279). New York, NY, USA: ACM. https://doi.com/10.1145/2851581.2892311

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI'94; pp. 72–78). New York, NY, USA: ACM. https://doi.com/10.1145/ 191666.191703

Piccolo, L. S. G., Mensio, M., & Alani, H. (2018, October). *Chasing the chatbots: Directions for interaction and design research.* Paper presented at CONVERSATIONS 2018, St. Petersburg, Russia. Retrieved from https://conversations2018.files.wordpress.com/ 2018/10/conversations_2018_paper_10_preprint1.pdf

Poliakoff, E., Beach, N., Best, R., Howard, T., & Gowen, E. (2013). Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands. *Perception*, *42*(9), 998–1000. https://doi.com/10.1068/p7569

Pollick, F. E. (2010). In search of the uncanny valley. In P. Daras & O. Mayora Ibarra (Eds.), *User Centric Media: First International Conference* (UCMedia; pp. 69–78). Berlin, Germany: Springer-Verlag. https://doi.com/10.1007/978-3-642-12630-7_8

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). Alexa is my new BFF: Social roles, user satisfaction, and personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI'17 EA; pp. 2853–2859). New York, NY, USA: ACM. https://doi.com/10.1145/3027063.3053246

Qu, C., Brinkman, W.-P., Ling, Y., Wiggers, P., & Heynderickx, I. (2014). Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior, 34*, 58–68. https://doi.com/10.1016/j.chb.2014.01.033

Schuetzler, R. M., Grimes, M., Giboney, J. S., & Buckman, J. (2014). Facilitating natural conversational agent interactions: Lessons from a deception experiment. In *Proceedings of the International Conference on Information Systems* (ICIS'14; online). Retrieved from http://aisel.aisnet.org/icis2014/proceedings/HCI/9/

Statista (2018). *Most popular mobile messaging apps worldwide as of October 2018, based on number of monthly active users*. Retrieved from https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/

Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015). Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI'15; pp. 3593–3602). New York, NY, USA: ACM. https://doi.com/10.1145/27021 23.2702415

Thies, I. M., Menon, N., Magapu, S., Subramony, M., & O'Neill, J. (2017). How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. In *Proceedings of the IFIP Conference on Human-Computer Interaction* (INTERACT 2017; pp. 441–459). Cham, Switzerland: Springer.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017, May). *Online human–bot interactions: Detection, estimation, and characterization.* Paper presented at the 11th International AAI Conference on Web and Social Media (ICWSM-17), Montreal, Canada. Retrieved from https://arxiv.org/abs/1703.03107

Vtyurina, A., Savenkov, D., Agichtein, E., & La Clarke, C. (2017). Exploring conversational search with humans, assistants, and wizards. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI'17 EA; pp. 2187–2193). New York, NY, USA: ACM. https://doi.com/10.1145/3027063.30 53175

Waddell, K. (2017, April 21). Chatbots have entered the uncanny valley. *The Atlantic* [online]. Retrieved from https://www.theatlantic.com/technology/archive/2017/04/uncanny-valley-digital-assistants/523806/

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior, 23*(5), 2538–2557. https://doi.com/10.1016/j.chb.2006.05.002

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45. https://doi.com/ 10.1145/365153.365168

Worswick, S. (2018, September 13). Mitsuku wins Loebner Prize 2018! [Web log post]. Retrieved from https://medium.com/pandorabots-blog/mitsuku-wins-loebner-prize-2018-3e8d98c5f2a7

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI'17; pp. 3506–3510). New York, NY, USA: ACM. https://doi.com/10.1145/30254 53.3025496

Yang, Y., Ma, X., & Fung, P. (2017). Perceived emotional intelligence in virtual agents. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI'17; pp. 2255–2262). New York, NY, USA: ACM. https://doi.com/10.1145/3027063.30 53163

Authors' Note

The work presented in this paper was conducted as part of the research project Human–Chatbot Interaction Design, supported by the Research Council of Norway through Grant No. 270940 of the IKTPLUSS programme.

All correspondence should be addressed to
Marita Skjuve
SINTEF
PB 124 Blindern, 0314 Oslo, Norway
marita.skjuve@sintef.no

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 15(1), February 2019, 55–78

EFFICIENCY 4.0 FOR INDUSTRY 4.0

Naveen Kumar
Department of Design
Indian Institute of Technology Delhi
New Delhi, India

Jyoti Kumar
Department of Design
Indian Institute of Technology Delhi
New Delhi, India

Abstract: *Operators of cyber-physical production system (CPPS) will use handheld interactive control panels for supervising production processes. These interactive and handheld control panels will present numerous sensors data. Processing of such complex data for task fulfilment in critical scenario will cause cognitive workload on operators. The prevalent industry efficiency measures so far have focused on human physical workload measures; the human cognitive load has not been considered in industrial practices. Smart factories in the near future will begin to integrate intelligent machines with complex communication network systems between human and machines. In this paper, we argue that the changing industrial scenario requires developing new measures of production efficiency. As a result, we propose a theoretical framework for a smart factory efficiency measure as a function of the combined efficiencies of the human cognitive system, smart machines, and their shared communication systems.*

Keywords: *Industry 4.0, HCI, human cognitive efficiency, production efficiency.*

DOI: https://doi.org/10.17011/ht/urn.201902201608

INTRODUCTION

With successive industrial revolutions, the workload of factory operators increasingly has shifted from physical exertion to cognitive workload due to the increase in complexity of information and the increasing use of information communication technology in factory automation (Wittenberg, 2015). In Industry 1.0, often a single operator was appointed to supervise the complete production process from the electromechanical dials attached locally to machines (Deane, 1965; Streans, 2013). The operator had to walk around the factory, going from machine to machine to gather information about the status of the machines and the production process. In Industry 2.0, it became possible to monitor the mass production process parameters from isolated control rooms. Large control panels with dedicated sections for machine parameters were monitored by multiple specialized operators (Astrom, 1985). In Industry 3.0, the current iteration in industrial evolution, computer-based manufacturing systems are integrated into the production operations and often involve multiple computers (i.e., the programmable logic controller & the supervisory control and data acquisition systems). In some operations, robots are used to supervise the production process remotely (Bailey & Wright, 2003; Briones, Bustamante, & Serna, 1994).

Industry 4.0, predicted as the fourth industrial revolution, will bring together information technology and factory automation (MacDougall, 2014). Industry 4.0 will use the Internet of things (IoT) in production systems to create a cyber-physical production system (CPPS; Lee, Bagheri, & Kao, 2015). The CPPS is envisioned as the core technology of Industry 4.0, which will comprise technologies such as wireless embedded network systems, IoT, and network cloud computing in a manufacturing setting (Sniderman, Mahto, & Cotteleer, 2016). CPPS is characterized by sharing production operation information over the Internet with multiple systems in the smart factory. Additionally, machines will be able to communicate production data with each other using embedded network systems. In Industry 4.0, the use of small, handheld mobile devices to control and monitor the production system will become prevalent. Finally, most factory data will be supervised through IoT-based human–computer interaction (HCI) systems. With the merging of information communication technologies (ICTs) with automation in a CPPS, production units will become smarter but also more complex at the visual, auditory, and other sensory levels (Wittenberg, 2015). The consecutive industrial revolutions are shown in Figure 1, as the ways humans and machines interact have evolved from Industry 1.0 to Industry 4.0 technology.

The increasing information complexity and its influence on production efficiency is a multidisciplinary challenge, where knowledge from cognitive psychology, ergonomics, operations management, communication technology, computer science, industrial design, HCI, manufacturing technology, instrumentation engineering, and others, must be addressed for the optimal design of control and display units in CPPSs. In this paper, we have drawn on concepts from multiple disciplines to develop a measure of industrial efficiency within this increasingly information-complex context.

Contemporary research literature has shown that the CPPS concept enables an ecosystem of cyber manufacturing in the factory environment (Lee et al., 2015). In the CPPS environment, smart machines process the production data through a wireless embedded network system to manufacture smart products (Li et al., 2017). This was demonstrated in a case study of a wireless network design prototype in small factory conducted by Wang, Wan,

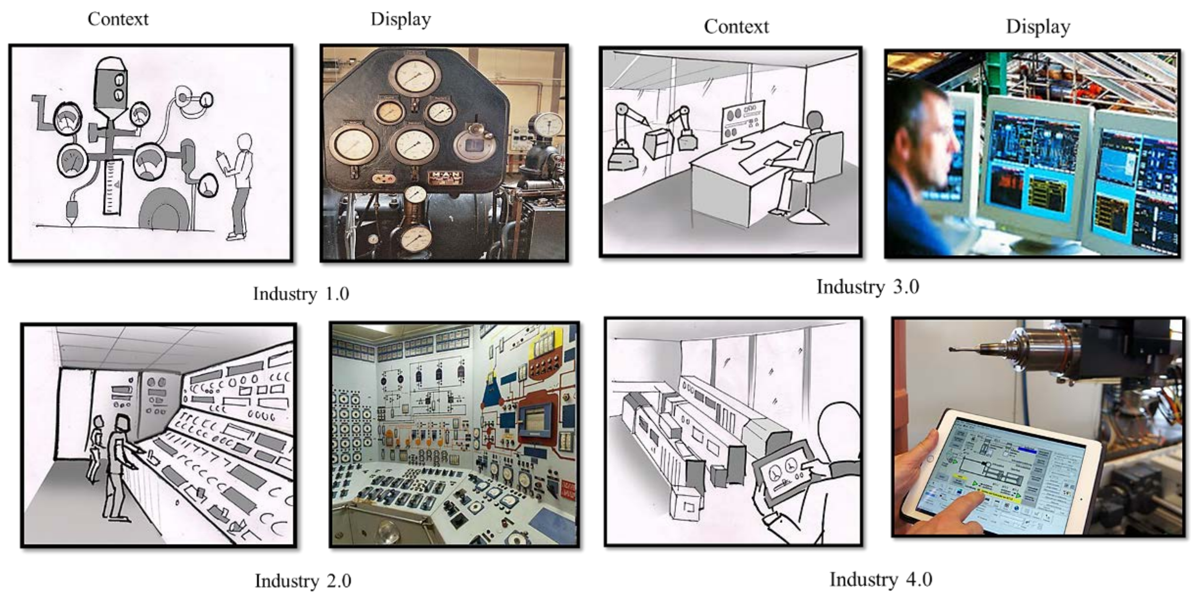


Figure 1. The significant changes in the process scenarios of human–machine interaction from Industry 1.0 to Industry 4.0.

Li, and Zhang (2016). As the manufacturing industry has slowly developed, the concept of Industry 4.0 (Posada et al., 2015), smart factories (MacDougall, 2014), networked manufacturing (Davis, J., Edgar, T., Porter, J., Bernaden, & Sarli, M., 2012), and other new frameworks (e.g., Wu, F. J., Kao, Y. F., & Tseng, Y. C., 2011; Zhang, D., Wan, J., Liu, Q., Guan, X., & Liang, X., 2012) have been proposed. Meanwhile, new ICTs, such as the industrial cloud (Wan, J., Zhang, D., Zhao, S., Yang, L., & Lloret, J., 2014), big data (Chen, M., Mao, S., & Liu, Y., 2014), wireless cloud networks (Wan et al., 2013), the industrial Internet of things (Lee et al., 2015), and high performance embedded systems have been introduced into the manufacturing processes to meet the demands for higher productivity.

Moreover, due to technological interventions in CPPSs, production information can be stored to and retrieved from networked clouds anywhere and anytime by human operators and as well as smart machines. Human operators can access production information through Internet-operated HCI devices. Smart machines can record sequences of human actions on machine interfaces to develop intelligent algorithms and, subsequently, reduce human workload in the production processes. Adaptive automation is the term used for this phenomenon (Kay, 2006; Villani et al., 2017).

Although CPPS machines are becoming more intelligent and may control and monitor the processes on their own by employing machine-learning algorithms (Lee et al., 2015), human interventions in production processes will still be required. Research in this area suggests that this monitoring of CPPSs will involve handheld mobile devices with rich interactive user interfaces (Meixner, Petersen, & Koessling, 2010) to monitor and operate the factory machines, often presenting complex information for real-time decisions (Villani et al., 2017; Zhong, Dai, Qu, Hu & Huang, 2013). This increased complexity in the data for humans in control operations will create increasing loads on the human information processing system; this information load in the brain is known as cognitive load (Sweller, 1988). Due to higher attentional requirements in the tasks, human working memory is likely to be severely strained with relevant and irrelevant data

(Brunken, Plass, & Leutner, 2003). Such data overload will increase the possibilities for human errors in tasks and a decrease in overall human performance. Thus, we posit here that the cognitive load caused by the design of these complex interfaces will become an important factor in the efficiency measure of the Industry 4.0 factories.

In current manufacturing processes, the efficiency of production units is measured by machines and manpower utilization functions (Subramaniam, Husin, Yusop, & Hamidon, 2009). While physical manpower played a central role in industrial efficiency of the first industrial revolution (Subramaniam et al., 2009), successive industrial revolutions have decreased the burden on human's physical workload while increasing the cognitive load due to an upward trend in information complexity and task complexity (Wittenberg, 2015). However, the efficiency measures prevalent in industry are still using manpower utilization as the central concept. Hence, modern industry needs to accommodate the cognitive workload in the efficiency measures (Subramaniam et al., 2009). In this paper, we propose a framework to measure the *human cognitive efficiency* (HCE) as a factor in overall industry efficiency.

Further, because ICTs in Industry 4.0 will use multiple communication protocols and network devices for making production processes functional across multiple shop floors, the centralized supervision of machines will become possible (MacDougall, 2014; Zhong et al., 2013). Therefore, we posit that the communication efficiency of network systems will increasingly affect the efficiency in future smart factories in Industry 4.0. To accommodate the network efficiency within the industry efficiency measure, we propose a *communication system efficiency* (CSE) component as part of the overall industry efficiency measure.

Moreover, as machines become endowed with increased information-processing capacity, the efficiency of the machines will require alternative considerations beyond the mere manual output-to-input ratio that is used presently in industry efficiency measures. Although machine efficiency measures as part of industrial production efficiency have been reported in literature (Daniel, 2006; Muchiri & Pintelon, 2008; Subramaniam et al., 2009), large data storage and fast processing capabilities of smart machines have not been factored into present measures. Smart machines in CPPSs are envisioned to have artificial intelligence algorithms that will optimize the production process operations automatically (Lee et al., 2015). Thus, we also propose in this paper the concept of *smart machine efficiency* (SME) as a factor in the overall production efficiency measure.

Thus, we employ the term *smart factory efficiency* (SFE) in this paper to represent the overall industrial efficiency measure. We propose that SFE be measured as a function of the above three measures, namely HCE, CSE, and SME. Table 1 provides a reference guide for the multiple terms and abbreviations used in is paper.

The aim of this paper is to highlight the need to reconsider the industrial efficiency measure within the changing context of the CPPS and then to propose a new framework for the efficiency measure. Thus, the contributions of this paper to the field are (a) the introduction of new industrial efficiency parameters, namely, HCE, CSE, and SFE, (b) the proposal of a mathematical relationship to measure cognitive load in industrial context, (c) a proposal of mathematical relationship to measure communication speed in industrial context, and (d) a proposal of a mathematical relationship to measure smart machine efficiency in industrial context.

In the following sections of this paper, we discuss prevalent production efficiency measures. We then propose a novel of smart factory efficiency measure, develop and discuss new

Table 1. Definitions of Significant Abbreviations in the Smart Factory Efficiency Framework.

Abbreviation	Term/Definition
CL	Cognitive load (on human mental processing)
CPPS	Cyber-physical production system
CSE	Communication system efficiency
HCE	Human cognitive efficiency
HE	Human efficiency
HPE	Human physical efficiency
K	Demographic and psychographic constant
OEE	Overall equipment effectiveness
SFE	Smart factory efficiency
SME	Smart machine efficiency
TE	Task efficiency

mathematical models to measure cognitive load, communication accuracy, and machine efficiency. We conclude the paper with a discussion of the implications of the proposed framework.

PREVALENT PRODUCTION EFFICIENCY MEASURES

At the beginning of industrialization, measurement of production efficiency was estimated by a simplistic output-to-input ratio. Later, more sophisticated measures were developed. For example, Cobb and Douglas (1928) proposed a mathematical relationship between inputs—namely, physical capital, labor, and the amount of output produced. The "Cobb-Douglas production function" is expressed as $Q (L, P) = A * L^{\beta} * P^{\alpha}$, where, Q (the quantity produced from the inputs) is a function of L (the amount of labor spent in the production), and P (the physical capital like machines, factories, computers, etc.). Further, in Cobb-Douglas production function, A is the measure of change in output that is not the result of the input (e.g., factors like the technology in a manufacturing setup), while α and β are the output elasticity, that is, the change in output that results from a change in either L or P values. For example, if $\alpha = 0.60$ and P is increased by 20% then output increases by 11%, given that other factors are constant.

Another measure, the overall equipment effectiveness (OEE), is quoted often in the literature (Muchiri & Pintelon, 2008; Nakajima, 1988). OEE has been used widely in industry as a production efficiency measurement. A direct relationship between the machine efficiency and manpower utilization in the OEE calculation also has been reported (Subramaniam et al., 2009). The term OEE was derived originally from the philosophy of the total productive maintenance concept, which states that the greater the level of factory automation in industry, the greater will be the cost reduction (Nakajima, 1988). OEE has been defined as a measure of total equipment performance (Muchiri & Pintelon, 2008) and it has been calculated as in Equations 1–5.

$$\text{OEE} = \text{AR} * \text{PR} * \text{QR} \tag{1}$$

$$\text{Availability Rate (AR)} = \frac{\text{Operating time} * 100}{\text{Loading time}} \tag{2}$$

$$\text{Operating time} = \text{Loading time} - \text{Down time} \quad (3)$$

$$\text{Performance Rate (PR)} = \left[\frac{\text{Machine cycle time* Actual output}}{\text{Operating time}}\right] \quad (4)$$

$$\text{Quality Rate (QR)} = \frac{\text{Total production–Defect amount}}{\text{Total Production}} * 100 \quad (5)$$

Although the OEE tool has been one of the most prevalent measures of production efficiency of manufacturing units, it has been criticized as well. Reports in the literature state that the calculation of OEE alone is not sufficient because no machine is isolated in the production unit (Oechsner et al., 2002). Scott and Pisa (1998) argued that the production process is a complex network of interactions among materials, machines, humans, and other processes; these interdependent factors have not been merged together to measure the production efficiency. Lee (2015) suggested a five-function framework for CPPS, involving smart communication, data-to-information conversion, cyber-level, cognition level, and configuration level. He has compared prevalent production efficiency measures with proposed Industry 4.0 production system efficiency factors and argued for improvement in OEE equations when applied to Industry 4.0. Similarly, Chen et al. (2018) reported a case study of the application of OEE in the Industry 4.0 context and argued for need to include more factors in the efficiency measure. Because the differentiating factor in Industry 4.0, as compared to Industry 3.0, is the use of information technology in machines that will increase the complexity of interactions (Gorecky, Schmitt, Loskyll, & Zühlke, 2014), we argue here for expanding the OEE to accommodate the changed environment of production. Additionally, there is a need to identify the factors that will affect the new measures in SFE.

PROPOSAL OF A FRAMEWORK FOR SMART FACTORY EFFICIENCY (SFE) MEASURE

The context of Industry 4.0 will shift the focus from production as an equipment-led activity to a smart factory activity that uses intelligent communication technology for industrial production processes. Up to now, the industrial production efficiency has been measured as OEE, which is calculated as a product of availability, performance, and quality as in Equations 1–5 (Muchiri & Pintelon, 2008; Nakajima, 1988). We are proposing a new framework, the SFE, to cover the newer dimensions that have emerged in production efficiency. The SFE is measured as a function of HCE, CSE, and SME, as expressed in Equation 6.

$$SFE= f(HCE, CSE, SME) \quad (6)$$

The next subsections present each of these three efficiency measures in detail. Then, following that, we provide a comparative discussion of OEE versus SFE.

Human Cognitive Efficiency (HCE)

Human efficiency in the industrial context can be an outcome of the combined cognitive and physical efficiencies, as in,

$$HE = f(HCE, HPE)$$

where HE is human efficiency, HCE is human cognitive efficiency, and HPE is human physical efficiency.

However, due to the adaptive automation movement in the design of industrial information systems (Kay, 2006), the human physical tasks are increasingly being allocated to the machines. Still, the HPE will require "human-in-the-loop" control models that will necessitate increased cognitive abilities with aid of intelligent human–machine interfaces (Romero, Bernus, Noran, Stahre, & Fast-Berglund, 2016). Smart factory supervision will demand higher cognitive workloads (Villani et al., 2017). Thus, the cognitive load will become a far more important consideration within the HE measurement in the Industry 4.0 context, as compared to the physical workload. Hence, in our research, HE has been proposed to be measured as a function of cognitive efficiency alone, as represented in Equation 7.

$$HE = f(HCE) \tag{7}$$

Paas and Van Merriënboer (1993) reported an attempt to measure "mental efficiency." They proposed an empirical formula for "performance and mental effort" as a measure of mental efficiency, as in Equation 8.

$$\text{Mental Efficiency} = \text{Mental Effort} - \text{Performance}/\sqrt{2} \tag{8}$$

As per Paas and Van Merriënboer (1993) in Equation 8, mental effort is the effort toward holding task information within one's cognitive capacity, whereas performance is one's success in the task. They had proposed this relationship based on the subjective self-assessed mental effort reported by problem solvers. However, in Paas & Van Merriënboer model, subjective self-reported ratings on performance and mental effort were gathered after task fulfillment. Therefore, although individual users can be asked for subjective rating, it will be difficult to model HE across the industry using "felt" mental effort. One of the reasons is that the mental effort in an industrial task can result from various factors, including the industrial scenario, task performance, the environment, time pressure, criticality, and so forth. Some researchers have argued that mental effort indicates some portion of the cognitive load that is imposed exclusively by the type of the task, the time on task, response time, errors, and the environmental demand (Cain, 2007; Kablan & Erden, 2008). Cognitive load thus becomes a better measure of HE then mental effort. Therefore, we propose a framework to measure HE through a cognitive load measure. We propose to define HCE as

$$HCE = TE/CL \tag{9}$$

where CL is cognitive load and TE is task efficiency, to be measured as

$$TE = \frac{\text{Total number of correct tasks within given time}}{\text{Total number of tasks}} \tag{10}$$

TE is proposed to be a ratio of the successful tasks versus the total number of tasks. This ratio is also indicative of HE, but we argue that unless it is factored by the total cognitive load that the user underwent in task fulfillment, the TE itself is not a clear measure of the HCE. Therefore, the CL factor has been included in Equation 9 for measurement of HCE.

As per Equation 9, when a task is completed with a lower CL investment, the HCE will be higher. Thus, Equation 9 paves a way to measure the HCE within the context of a given design. As a result, the design of a control panel should enable the user to achieve the same task outcomes with a lower CL, thus making it a cognitively more efficient design. This also can indicate that the person has utilized fewer mental resources and therefore the operator has been more efficient. This approach to the HE measures has many implications. For example, the design quality judgment and personal evaluation of operators will become possible with the present framework. However, Equation 9 needs further detail in order to make such assessments possible. Thus, we propose that CL be measured as

$$CL = K * I_P/t \tag{11}$$

where, K represents a constant that reflects user's demographic and psychographic factors such as age, stress level, competency, and so on. I_P is the information to be processed during task (unit: meaningful chunks) and t signifies the time for the task duration (unit: minutes).

Equation 11 is proposed to measure CL in terms of information processed (I_P) by users' during task fulfillment in the CPPS context where, for instance, they need to make sense of complex shop floor data, decide on a course of actions, and then complete those actions through the HCI modules that may be stationary or mobile, remote or localized. The information processed per task is proposed to be measured in the unit of "number of meaningful chunks of information" because CL theory emphasizes the constraints on the human cognitive process due to limits on working memory capacity. This is in line with previous research, where units of meaningful chunks of information has been noted as the measurement for working memory capacity (Atkinson & Shiffrin, 1968). Miller reported that a human brain can store in the working memory a maximum of nine meaningful chunks of information at a time, on average (Miller, 1956). Saaty and Ozdemir (2003) added that, when a person is making preference judgments on pairs of elements in a group, the number of elements in the group should be no more than seven. However, recent studies have suggested that the working memory capacity is reduced to three to five meaningful items (Cowan, 2016). Thus, based on the literature (Cowan, 2016; Miller, 1956), meaningful chunks of information (I_P) has been proposed as a measure for CL.

The unitless constant K is proposed to serve as a personal demographic and psychographic factor that can influence the user's cognitive performance and can be attributed to the elements such as the age, stress level, and task competency level of the user. Studies have reported the influence of age on cognitive performance. For example, Caplan and Waters (2005) reported slower speed in judgment and paragraph comprehension tasks by elderly study participants. Salthouse, Babcock, and Shaw (1991) reported that older adults were less accurate than young adults in mathematical tasks and hypothesized it to be caused by decreased working memory capacity. Similarly, stress has been reported to influence memory performance (Lupiena & Lepage, 2001). Thus, the constant K in Equation 11 captures these personal constructs affecting the CL. Although prevalent literature reveals the influence of age, stress, and competency on the experienced CL, little evidence is available on the exact proportion of effect of these factors on CL. Therefore, we do not address in this paper the possible values or methods to calculate a constant K.

We argue here that the quality of CPPS-based HCI designs can be measured in terms of the CL. Higher the complexity in task and information, more improvement in the interface design would be require. As per Equation 11, the effort to redesign user interfaces can be directed toward increasing the semantic quality of interfaces. We coined the term *semantic quality* here to address the lack of an appropriate term to describe the ease with which the user of an interface can perceive meaningful information within the context of a task. Because working memory capacity is measured in terms of meaningful chunks of information, any improvement in the semantic quality of the user interface design will reduce the number of chunks of information required to be held in working memory for task fulfillment, thus reducing the redundant information and/or increased the time needed to comprehend a given set of information. The guidelines in Equation 11 will aid designers of HCI systems in increasing HE in the context of a CPPS.

We note that the proposed HCE is a measure of total mental effort spent in course of completing successful tasks. It is a system measure and not an individual capacity measure. This construct measures the cognitive efficiency of the (generic) human in a cyber-physical production system due to the design quality of the system.

Further, the literature identifies CL as comprising three types (see Paas, Renkl, & Sweller, 2003), namely, extraneous CL (the load caused due to adding irrelevant information in the task), intrinsic CL (the inherent level of task complexity), and germane CL (the load required to develop a new schema). We therefore propose that, in order to investigate the specific causes of CL in specific tasks in a CPPS, the measurement of HCE should zoom into the type of CL required of the user for each task. Thus, we posit here that the CL on a user can be calculated as a function of the three types of CL, as in Equation 12:

$$\text{CL} = f(\text{Intrinsic CL, Extraneous CL, \& Germane CL}) \tag{12}$$

Intrinsic CL has been defined as a function of task complexity and can be calculated by the total information to be processed to complete the task (Paas et al., 2003). For example, the multiplication of two single-digit numbers requires a lower intrinsic CL as compared to multiplying two three-digit numbers (Kumar & Kumar, 2016). Therefore, intrinsic CL can be calculated by estimating the complexity of the task, indicated by the number of steps involved in the task. Intrinsic CL can be written as Equation 13:

$$\begin{aligned}\text{Intrinsic CL} = \text{CL (Task Complexity)} = f(&\text{no. of steps in task, amount of}\\ &\text{information to be processed in each step}\\ &\text{of task, frequency of task, etc.})\end{aligned} \tag{13}$$

Further, in cognitive load theory, extraneous CL is defined as CL caused due to presentation of irrelevant information in the task to the user (Paas et al., 2003). Thus, we propose here that, in the context of a CPPS, extraneous CL be measured as CL caused by the difficulty in the semantic interface that results from poor design quality in the user interface, specifically presenting irrelevant information, low-information graphics, incomprehensible textual content and metaphors, inconsistent layout, gratuitous use of colors, incomprehensible feedback, undifferentiable tactile responses, and so on. Hence, extraneous CL can be represented as in Equation 14:

$$\begin{aligned}\text{Extraneous CL} = \text{CL (Semantic interface)} = f(&\text{consistency of layout, legibility of}\\ &\text{text and information graphics, ease and}\\ &\text{efficiency of navigation, contrast between}\\ &\text{foreground \& background color schemes}\\ &\text{choice of displays, buttons, menus, etc.})\end{aligned} \tag{14}$$

Germane CL is caused by the person's need to create a new schema in the brain because of interacting with the new task or design of the system (Paas et al., 2003). In context of CPPS-based HCI systems, we propose that the germane CL be measured by *task novelty*, which is the effort exerted by the user to learn the new way of task fulfillment. Germane CL be measured as in Equation 15:

$$\text{Germane CL} = \text{CL(Task Novelty)} = f(\text{difference between the old and new layout of designs/task steps/input methods, display methods, etc.}) \tag{15}$$

Task novelty can be understood as the effort expended by the user to create a new mental model (Johnson-Liard, 2005), or a new schema of the task within the user, that is, creating new internal guidance to complete the task. The new schema is stored in the long-term memory and can be fetched by the user into the short-term memory during task completion. A robust and easily comprehensible schema of the task resulting from the design of the interface will thus help in the HE factors. However, a change in technology and the user interface designs means such schemas will be created by repeated task fulfillment activities rather than mastered in one go. Thus, as depicted in Equation 15, germane CL can be measured as a function of difference between the old and new mental models (Johnson-Laird, 2005) presented through layout designs, steps in tasks, input methods, display methods, and so forth.

The purpose of zooming into the causes of CL challenges at the level of specific HCI design elements by using these equations is to aid the systems designer in identifying possible new options to optimize the cognitive costs of tasks design. Moreover, the HCI designer can refer to prior research on human cognitive capacity (e.g., Cowan, 2016; Miller, 1956) to estimate the optimum level of CL requirements by distributing the information of tasks and time durations to assist the operators. Thus, the HCE measure using these equations can evaluate not only the status of the system, but it also can guide the designers of a CPPS to improve the system during the design process.

This section has proposed equations to measure a technology system's HCE as a factor of three types of CL, namely, intrinsic, extraneous, and germane. We also discussed the factors influencing these types of CL in the CPPS context. Next, the CSE is discussed.

Communication System Efficiency (CSE)

In CPPSs, smart machines communicate wirelessly with other machines and humans. These smart machines also have the capacity to process large amounts of information in a very short time. A CPPS will reduce the need for human intervention, but human activity typically will be required to monitor and make critical decisions. Efficiency in a CPPS's communication system also will depend upon the speed and accuracy of information as it travels machine to machine, machine to human, and human to human. Developing a CSE measure requires a discussion of the context of a communication system within Industry 4.0

Complicated usage of wireless communication is a given in Industry 4.0. Wireless communication technologies—such as Bluetooth, machine-to-machine communication, radio frequency communications, Wi-Fi, Zigbee—will be used over a secured Internet protocol (Da, We, & Li, 2014). As a continuation of electronic communications started in Industry 3.0, Industry 4.0 will increase the types of communication protocols to manage the types and amount of information. We posit, therefore, that the CSE is a measure of how quickly the correct information will reach to the desired nodes in the system. Equation 16 presents CSE as a function of speed and accuracy of the information.

$$\text{CSE} = f(\text{Speed of information, accuracy of the information}) \qquad (16)$$

In order to argue for the development of Equation 16, we provide a likely picture of the CPPS environment in Figure 2. A typical CPPS system will consist of multiple smart machines within each production unit that could be spread across several shop floors and/or at multiple locations. Figure 3 presents a schematic diagram of a small section of the CPPS communication system. From this, one can observe that information is flowing from machines (M1 to M4) to the human operator (H1) through network nodes (N1 to N4).

Figure 3 shows a section of the CPPS communication nodes where information from various machines (M1, M2, M3, & M4) flows into the network nodes (N1, N2, N3, N4) and is finally received by the human (H1) for factory production process supervision and making decisions. A factory operator can supervise multiple process parameters of the shop floor through HCI devices from remote locations. For example, a hypothetical user interface of an HCI device has been developed to present the CPPS scenario (shown in Figure 4). A factory

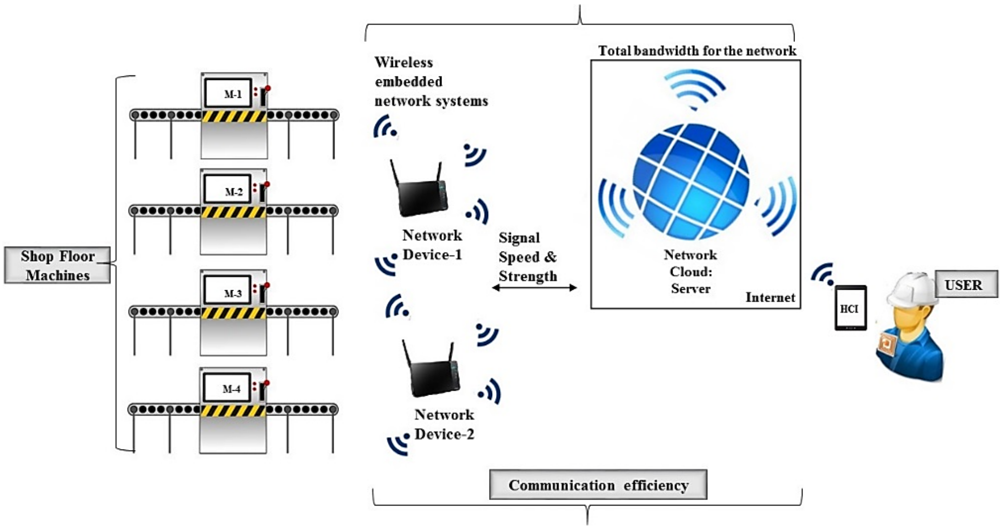


Figure 2. A human and machine communication scenario in a cyber-physical production system (CPPS) shop floor.

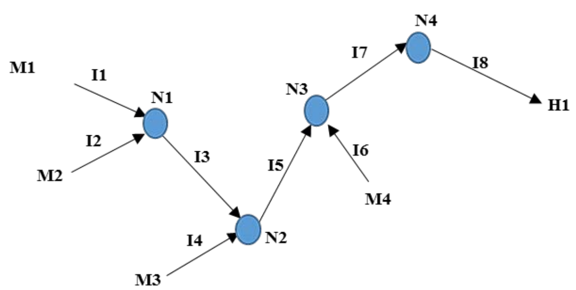


Figure 3. An example of machines-to-human communication via network nodes (N1–N4). In this schematic, I = information, N = network nodes, M = any variation of machine, and H = the human operator.

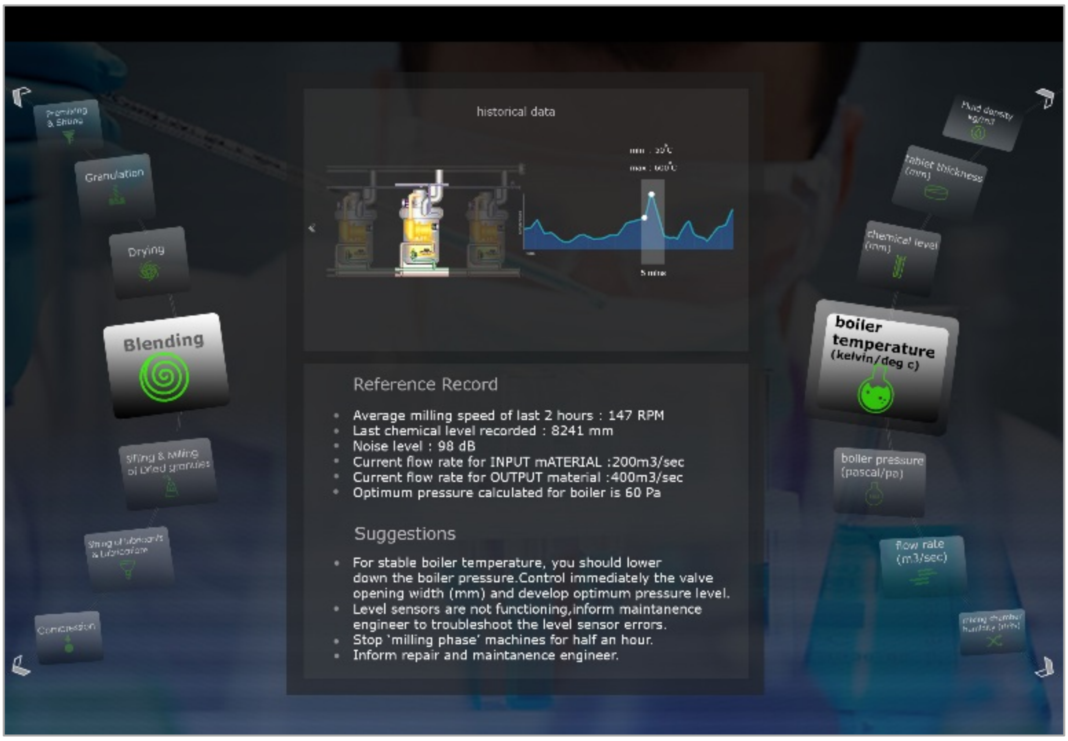


Figure 4. A hypothetical user interface of an HCI device.

operator (H1 in Figure 3) of a pharmaceutical manufacturing company can supervise the entire production process in real-time through this HCI device.

In Figure 4, a hypothetical user interface shows a monitoring screen with various production sections (blending, grinding, milling, etc.) of a CPPS. Process parameters also are represented visually in the user interface. Suppose the operator (H1 in Figure 3) needs the boiler temperature information available as I1 from machine M1; this information travels to the node N1. The total information received by the node N1 is I1+I2 (I1 & I2 are two process parameters, as there is another machine M2 connected with node N1). Similarly, other machines in the CPPS can send information (I3, I4, I5, etc.) to H1 through various other nodes. In this network of communications, the nodes are intermediaries between the human and the machine(s). We posit here that the summation of the individual communication efficiencies between the nodes will affect the SFE of the CPPS.

The efficiency of communication between the nodes can be measured as function of the speed of communication between nodes and the accuracy of the information communicated as depicted in Equation 16. Further, we propose that the speed of communication between nodes will be a function of bandwidth between the nodes, amount of information to travel, type of communication protocol, processing time, relay time, signal strength, number of nodes in a path, and so on, as presented in the Equation 17.

$$\text{Speed of information between nodes} = f(\text{bandwidth between the nodes, amount of information, type of communication protocol, processing time, relay time, signal strength, no. of nodes in a path, etc.}) \tag{17}$$

Similarly, the accuracy of information measurement is a function of loss of data packets between nodes, technical failure of nodes, and duration of waiting time of the information because of bandwidth congestion, as presented in Equation 18.

$$\text{Accuracy of the information between nodes} = f(\text{loss of data packets, technical failure of nodes, delay in information delivery}) \quad (18)$$

In the scenario depicted in Figures 3 and 4, a loss of data due to situations such as low signal strength, signal drop, and so on, will provide inaccurate or incomplete information at the user's end. Similarly, a delay in information delivery at the interface (Figure 4) will exhibit inaccurate, incomplete, or outdated monitoring data to the operator.

The combination of speed and accuracy will provide a measure of communication efficiency in CPPS. The result will indicate the optimum transfer of information from machines (M_1 to M_n) through nodes (N_j to N_k) to reach human operators (H_1 to H_n) such that the interaction of operators with HCI interfaces, like in Figure 4, will receive timely and accurate data and act appropriately. Summing up the CSE of the overall different nodes in CPPS, we propose Equation 19 to measure the overall efficiency as a summation of the speed across the nodes.

$$S(N) = \sum_{i} \sum_{j}^{k} \mathrm{Ii}(\mathrm{Nj} - \mathrm{k})/\mathrm{t} \quad (19)$$

The speed of the information in CPPS communication network is represented as S(N). Integers i, j, & k represent the number of information packets, node numbers, and total numbers of nodes, respectively. The i is the information amount to travel between the nodes, t is the total information travel time, and N is a node in the network.

Smart Machine Efficiency (SME)

Measures of efficiency in machines within the industry have been reported in the literature (e.g., Muchiri & Pintelon, 2008; Nakajima, 1988). However, as machines become smarter with capability of information processing and programmed logic within them, there is need to reassess at the machine efficiency measure. As a continuation of the arguments presented in the Introduction, we propose here that the measure of efficiency of the smart machines in a CPPS includes factors such as total time to produce all the product, total lost time in defective product manufacturing, total manufacturing time, machine information processing time, machine run time, machine delay time, machine errors, and so forth. We propose that measurement of SME as Equation 20:

$$SME = [(T_P - T_D)]/ M_T \quad (20)$$

where T_P represents the total time to produce products, T_D means time lost in defective product production, and M_T is total manufacturing time. Further, total manufacturing time (M_T) is proposed to be calculated as

$$M_T = M_R + D_T + M_W + M_P \quad (21)$$

where, M_R designates machine run time, D_T represents delay time due to technical failures (errors), M_W manual work time before and after the start of machine, and M_P the machine's information processing time.

The SME measured in Equation 20 is proposed to ascertain the percentage of time spent per non-defective product. As in Equation 21, the total time calculation is envisioned to consider all the various stages in the machine operation. For example, M_P indicates the time spent in information processing by a smart machine. These types of processing inefficiencies could be due to hardware or software design of the smart machines. Although currently this time is negligible and hence often is discarded in calculations, as machines become better integrated with computing capabilities in Industry 4.0, we argue for this computing time also to be considered in the SME measure. Similarly, D_T measures the delay due to technical failure in mechanical parts of the machine, electronic hardware, malwares in the software, power failures, and so on. The M_W measures the delay in machine operation due to the machine operators' pre- and postproduction planning efforts.

Additionally, we argue that, despite the automation, the CPPS efficiency also will depend upon smartness of the smart machine operators. Any delay caused by programming the information into the smart machines before production starts also will need to be included when designing the CPPS. Hence, the training of operators, the design of machine programming that leads to the need for more manual time, the intuitive design of user interfaces of the smart machines, and so forth, will contribute to the M_W.

To sum up, we discussed in this section that the prevalent machine efficiency measures are dependent upon manpower and machine run-time calculations, such as the ratio of operating time upon loading time as in Equations 2 and 3 (Nakajima, 1988; Subramaniam et al., 2009). Therefore, we have argued for the need to include these factors due to the smartness of new machines that will impact information processing time, delay time, manual operator time, and more within the SME measure.

Comparison Between Smart Factory Efficiency (SFE) and Overall Equipment Effectiveness (OEE)

Along the lines of prevalent practice of measuring the OEE (Muchiri & Pintelon, 2008; Nakajima, 1988), an overall SFE has been proposed here as a product of HCE, CSE, and SME, as in Equation 22:

$$SFE = HCE * CSE * SME \tag{22}$$

SFE was developed from the OEE measure as depicted in Equations 1–5. Equation 1 is presented again here for a comparative discussion of SFE and OEE measures:

$$OEE = AR * PR * QR$$

where, AR designates the availability rate, PR the performance rate, and QR the quality rate.
In OEE, the overall production efficiency measure depends upon manpower and machine run-time calculations, meaning the ratio of the operating time upon loading time, as discussed in Equations 2 and 3 (Nakajima, 1988). Manpower utilization has been an important factor in prevalent efficiency measures (Subramaniam et al., 2009). Equation 22, however, shows the shift in the focus of the man–physical power calculation from physical load to cognitive load, in line with the tenets of Industry 4.0. Further, due to the improving smartness of current and future machines, the efficiency measure has been accommodated in the SME calculations, and the increase in information load due to smart factories has been accommodated in the HCE

measure. While the HCE reflects human cognitive performance, the SME integrates the total manufacturing time with the time to produce quality products. Thus, the SME is a modified representation of the quality rate of the OEE measure depicted in Equation 5. Similarly, the OEE measure uses performance and quality factors that have been accommodated in HCE and SME measures in the SFE framework.

A CASE STUDY FOR HUMAN COGNITIVE EFFICIENCY (HCE) MEASUREMENT

In order to identify the HCE (Equation 9), we provided measures for the TE (Equation 10) and the CL (Equation 11). To validate the empirical equations of HCE, we undertook a brief experiment. For the experiment, an interactive touch-based control panel application was developed and installed onto a tablet device and then tested on single participant with 10 trials over consecutive days.

The participant was given a controlling task and the scenario was explained. The task was to control the process variable when it reaches above the critical values using the knobs provided on the test control panel. The total time for a complete session was five minutes, and the participant performed a total 16 controlling actions using the knobs on the control panel. When process variable value rose above the critical limit, the participant was to respond via the controlling knob. Response times of and the number of errors by the participant were recorded automatically by the control panel application.

The test participant, aged 24 years, possessed a fare knowledge of the dials and knobs needed to perform the experiment well. To ensure the collection of reliable response time and task error from the participant, we inquired into his medical history of eyesight and head injury; the participant had no medical issues that would negatively impact the experiment.

The control panel was designed based on a hypothetical pharmaceutical manufacturing company in the context of Industry 4.0. The control panel design had four display dials, namely boiler temperature, boiler pressure, station voltage, and motor speed, as well as four controlling knobs, namely steam-flow rate, valve-opening width, station load, and time. The testing facilitator provided the control panel tablet to the participant and then left the room. From that point onward in the experiment, the scenario was narrated by the control panel application software. The audio message provided by the machine began when the operator received a call from the machine:

> *Hello Sir, I am machine number #627. There is an urgent action required in plant operation.*
>
> *In the blending phase, level sensors have stopped working. System variables such as boiler temperature, pressure, motor speed, and station voltage have become unstable. I have tried to set all the parameters in calculated optimum range but could not succeed. Next, I am sending you the overall machine control of the blending phase. Try to keep all the display parameters in a stable range and give your response to critical conditions only. Click END BUTTON to receive the machine control.*

Figures 5 to 9 show the screens to the participant interacted with during the tasks. Figure 5 shows what the participant saw when receiving the call from the machine reporting the situation in the production process at the smart factory. With the presentation of Figure 6, the decision-making tasks (such as overall control) were given to the participant with the help of an audio message. Figure 7 shows that two parameters of the blending phase have reached a critical condition and the participant must control both the parameters one by one. Before that response, however, the participant must consider the machine-generated suggestions that are shown in Figures 8 and 9. Figure 8 presents the boiler temperature status and the suggestion generated by the machine. Similarly, Figure 9 provides boiler pressure information. The information provided the participant by Figures 8 and 9 established the parameters that demanded immediate controlling task from the participant.

In order to check the TE and CL of the operator, we created just one critical scenario as an example. However, a real-world scenario could involve many operational tasks and for different scenarios. For example, what if an error occurred in multiple phases (e.g., blending, drying) and with multiple process variables (e.g., flow, humidity, blower speed)? In such a situation, the way information of the multiple process variable is presented could become one of the factors for an operator to respond slower or faster on the knobs, which increases the operator's CL and affects the TE of the operator. Hence, considering the HCE as a measurement in SFE is important in task success in an Industry 4.0 environment.

During the trials, the participant sat on an isolated closed room and performed the task on an 8-inch tablet hosting the interactive control panel (see Figure 10). Both task response time and task errors were recorded automatically by the control panel application software during the task execution. Task response time is the amount of time expended by the participant while controlling the critical values of the display parameters. Regarding task errors, two types were recorded during the test. The first type of error was the wrong attempt, when the participant

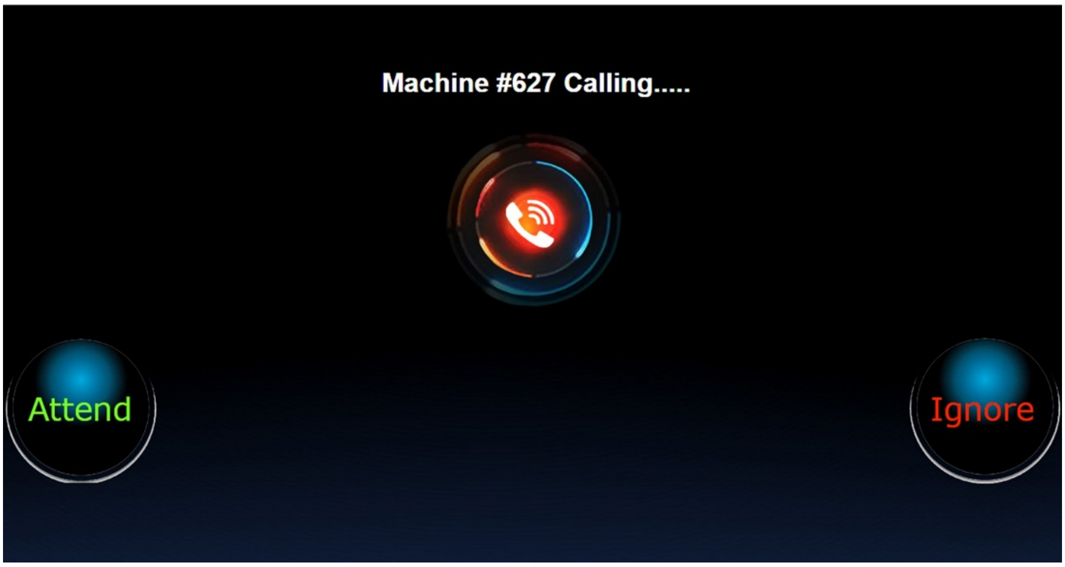


Figure 5. The screen demonstrates what the test participant saw when the machine called the operator to report factory critical condition.

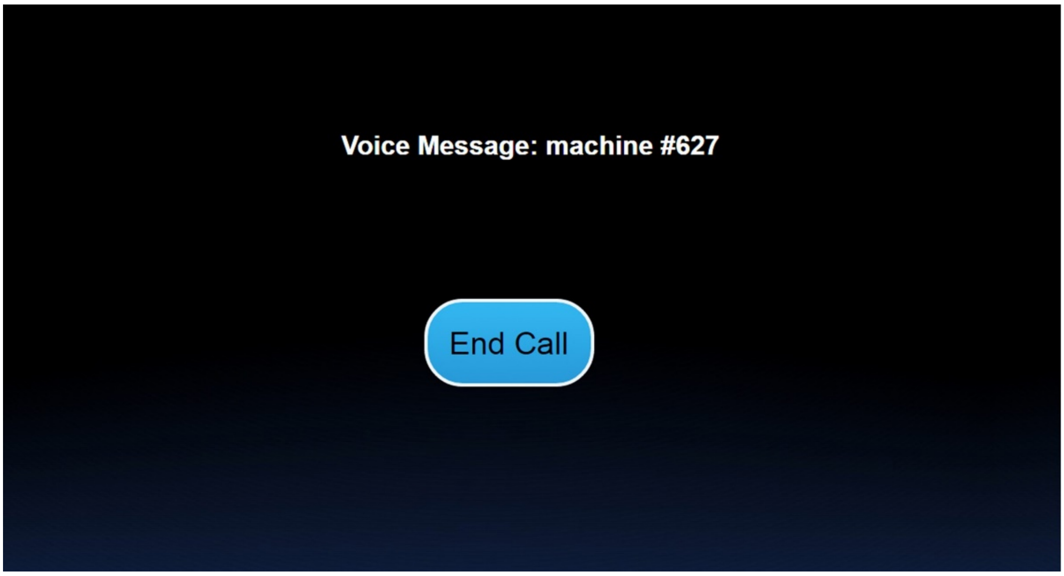


Figure 6. This screen appeared after the machine reported the problem to the operation and established a means for handing over factory machine control to the operator.

touched the wrong control slider. The second type was the missed chance that occurred when the participant did not respond within 10 seconds of the display of the critical value. A total error count was calculated by summing both types of errors.

Figure 7. The test operator saw this main screen that displayed the critical conditions for the two parameters.

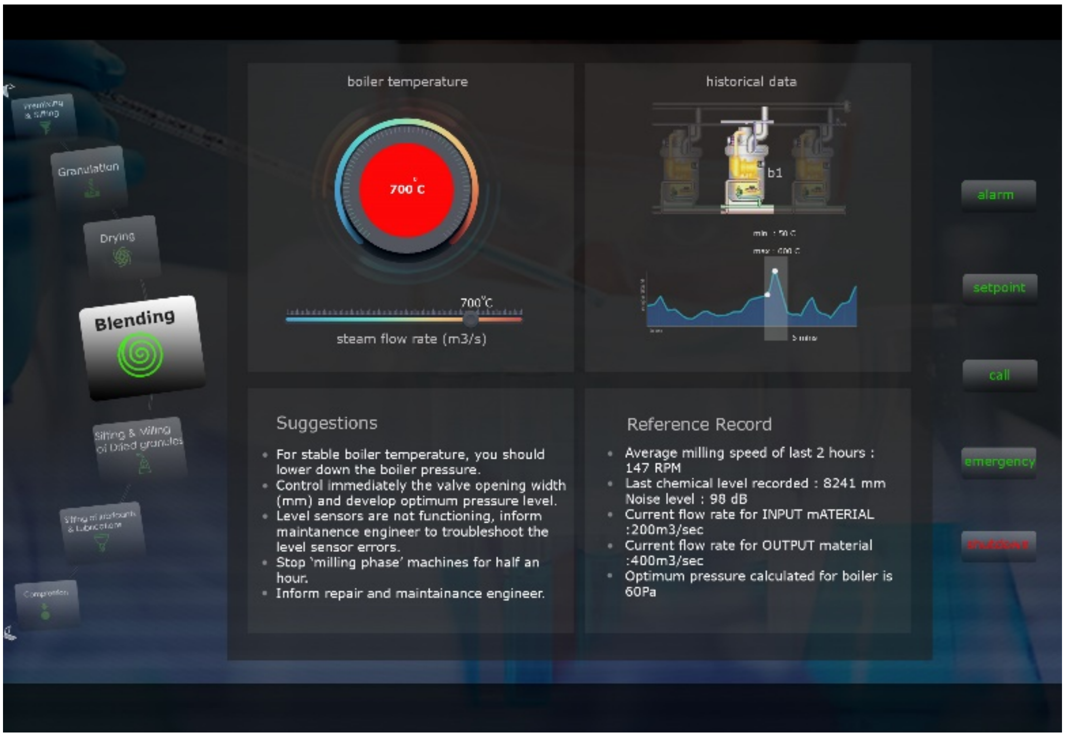


Figure 8. This control panel screen showed the operator the machine-generated status and suggestions for the boiler temperature parameter.

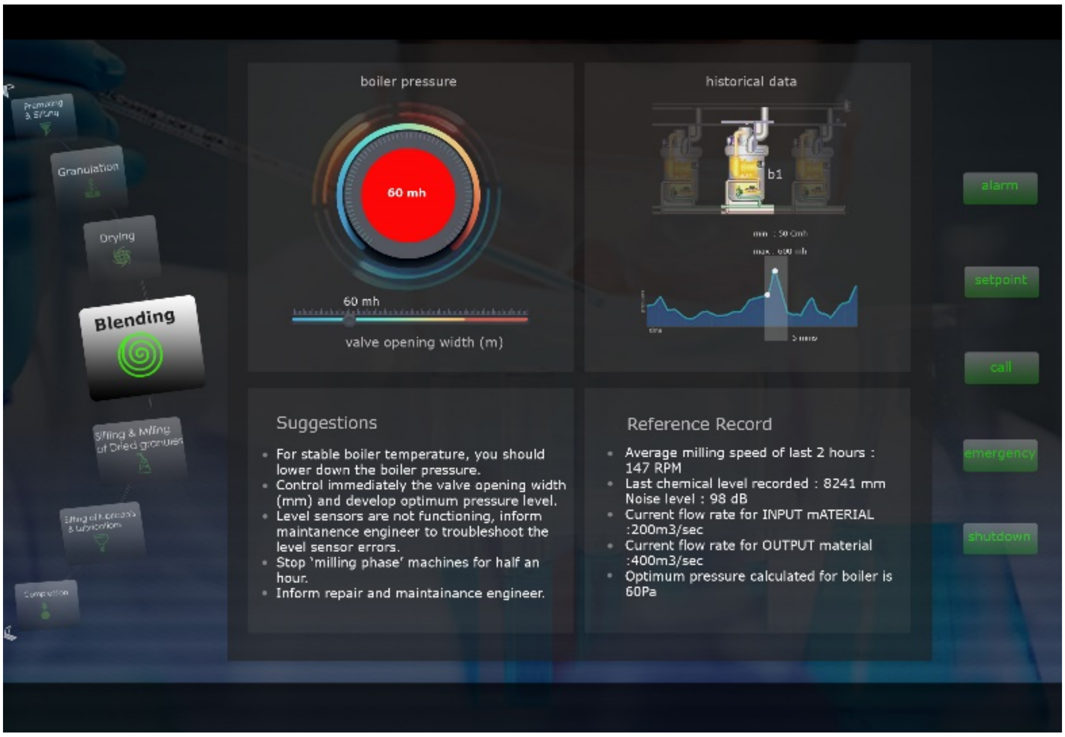


Figure 9. This control panel screen provided the machine-generated status and suggestions for the second critical condition: the boiler pressure parameter.

Figure 10. In line with the experimental setup, the participant used a touch-based control panel to perform each task in all 10 trials.

Based on the task errors, we were able to calculate the total number of correct tasks within a given time as compared to the total number of given tasks. With this information, we calculated the TE, which is presented in Table 2.

According to Equation 9, HCE = TE/CL. As per Equation 10, TE was being measured in the test case scenario. In our test, the average TE for a single user (the operator of the smart factory) with 10 trials was 0.8187 (i.e., TE = 81.87 %). Moreover, TE is directly proportional with HCE. If TE increases, HCE will also increase. TE has a direct relation with HCE.

From this result, we concluded that the TE of a factory operator can be measured in a critical Industry 4.0 scenario. The TE measurement can assist control panel designers to design efficient control panels for specific Industry 4.0 tasks and scenarios. However, based on Equation 11, the CL (CL=K*Ip/t) cannot be formally measured because researchers still must discover the way to calculate the value of the constant K in the equation.

While research has shown that human cognitive performance is influenced by age (Caplan & Waters, 2005), stress (Lupiena & Lepage, 2001), and competency level (Salthouse et al., 1991), the literature offers little regarding the proportionate effect of these factors on CL. Thus, this

Table 2. Task Efficiency Calculation as the Comparison of the Total Number of Correct Tasks Divided by the Total Number of Tasks over the 10 Trials.

No. of control panel trials	Total no. of correct tasks (T1)	Total no. of tasks (T2)	Task Efficiency TE=(T1/T2)
1	14	16	0.875
2	16	16	1
3	7	16	0.4375
4	9	16	0.5625
5	14	16	0.875
6	14	16	0.875
7	14	16	0.875
8	15	16	0.9375
9	12	16	0.75
10	16	16	1

paper has not proposed an exact method of calculation of K. However, this inability does not affect the relevance of the proposed efficiency measure in that efficiency is a relative concept where increases or decreases need to be measured. Even without an exact value for K,' the equation is still useful. Further research to calculate K is necessary to strengthen the means to measure CL and overall HCE.

CONCLUSIONS

The current efficiency measurement practices in Industry 3.0 lacks the consideration of human cognitive load even though workers' cognitive load is on a rise due to the increase in information complexity. Prior industry frameworks to measure production efficiency are not practically useful at this point in industrial evolution and certainly will not be able to do justice to the changed scenario of production and productivity of Industry 4.0. Hence, new theoretical models will be needed.

Moreover, although cognitive load theories, production efficiency theories, and communication theories have outlined the need to reconsider the production processes as a sustainable and profitable human activity, the theoretical frameworks to implement these theories adequately have yet not evolved. However, we have proposed a theoretical measure for SFE comprising three functions, namely, HCE, CSE, and SME (see Table 1 for abbreviation definitions). We have positioned the need for a new theoretical framework in context of the Industry 4.0. This rethinking of the industrial context is necessary because the Industry 4.0 environment will place an increased CL on the humans, reflect increasing smartness in new machines, and require increasing levels of communication between machines. These conditions form the argument underlying our proposal for the new framework.

We observed one common trait in these three functions (HCE, CSE, and SME): "information." The role of information in current and emerging industrial realities affects not only the technical side of the equation but, most importantly, the human operator. If information effects the human cognitive system due to increased information complexity, task complexity, and critical scenarios, then human cognitive efficiency will suffer. If an overload of information across multiple nodes (i.e., exchange of data from servers to different nodes) creates malfunctions or errors in the communication protocol—or is not presented clearly to the human operator, then communication system efficiency will suffer. Similarly, if irrelevant information or noise in the machine's interrupts or delays data processing in any, then smart machine efficiency will suffer. Therefore, we advocate for the need to look anew at production efficiency and consider information as a central component for the measurement of smart factory efficiency. The framework presented in this paper is our attempt to open discussion and re-vision efficiency measurement in current and future work environments.

Every research process is constrained by limitations. The most significant limitation of this work has been that the proposed theoretical framework does not quantify the individual contributions on the three efficiency measures toward an overall SFE measure. We are unable to do so at this time because the research field has not yet been able to define a value for a constant K, and so the proposed Equation 11 measurement for CL is not yet fully developed. This will change in the future as researchers conduct experimental studies to measure K.

What could be considered another weakness of the proposed framework is that the role of physical workload has been removed completely from the mathematical models based on the assumption of full automation of the industrial environments in Industry 4.0. However, we do acknowledge that some industrial contexts will continue where human physical labor will be used to some extent in the production process. Therefore, we see that this framework will require improvements.

For the proposed framework to gain popularity in mainstream industrial efficiency measures, further research work is required. The evaluation of the design quality of HCI-based control systems using the CL measures and HCE framework can be one area of future research studies. Similarly, aspects of CSE and SME can be evaluated by simulations of a futuristic CPPS to assess effectiveness of the proposed system designs.

IMPLICATIONS FOR RESEARCH OR APPLICATION

As technology advances on multiple levels, the impact on industrial practices will change as well. Current manufacturing and other factory environments and processes have already reflected increased reliance on computers and smart machinery. In years to come, with the emergence of Industry 4.0, machines will be the primary drivers of factory work, with humans serving in the troubleshooting and oversight roles. Our research looks toward that future in presenting a framework specifically envisioned to assist industry managers and workers in establishing and measuring the efficiency of their operations—specifically regarding their workers overseeing the operations.

From a practical perspective, our framework allows industrial technology designers and engineers to look differently at what measurements are useful in creating systems that are efficient not only from the technological perspective but also from the human perspective. In particular, understanding how technology advances may challenge the human workers' cognitive load (which in turn affects their efficiency) can lead to optimal HCI design (with sufficient technology communication efficiency and smartness) and, especially, better technology interfaces for human interaction with the machines. Our framework also provides a practical means—through read-to-use formulae—to evaluate efficiency at the various levels of the future factory environment.

Apart from the practical contributions, our research contributes to the future development of industrial efficiency measurement by way of our bringing together interdisciplinary knowledge to create a framework. The case study of human cognitive efficiency measurement provides groundwork upon which future researchers can continue to test and develop the framework. The proposed mathematical equations open a scope for researchers to investigate more fully the human factors, telecommunication technology, and embedded systems in order to make information technology more efficient.

REFERENCES

Astrom, K. (1985). Process control: Past, present and future. *IEEE Control System Magazine, 5*(3), 3–10.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation, 2,* 89–195.

Bailey, D., & Wright, E. (2003). *Practical SCADA for industry*. Burlington, MA, USA: Elsevier.

Briones, L., Bustamante, P., & Serna, M. A. (1994). Wall-climbing robot for inspection in nuclear power plants. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 1409–1414). San Diego, CA, USA: IEEE.

Brunken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 53–61.

Cain, B. (2007). *A review of the mental workload literature* (NATO RTO report). Toronto, Ontario, Canada: Defense Research and Development.

Caplan, D., & Waters, G. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory*, *13*(3-4), 403–413.

Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., & Yin, B. (2018). Smart factory of Industry 4.0: Key technologies, application case, and challenges. *IEEE Access, 6*, 6505–6519.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *The American Economic Review*, *18*(1), 139–165.

Cowan, N. (2016). *Working memory capacity*. New York, NY, USA: Psychology Press.

Da, X. L., We, H., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transaction on Industrial Informatics, 10*(4), 2233–2243.

Daniel, D. (2006). Overall input efficiency and total equipment efficiency. *IEEE Transaction on Semiconductor Manufacturing, 19*(4), 496–501.

Davis, J., Edgar, T., Porter, J., Bernaden, J., & Sarli, M. (2012). Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Computers & Chemical Engineering, 47*, 145–156.

Deane, P. (1965). *The first industrial revolution.* Cambridge, UK: Cambridge University Press.

Gorecky, D., Schmitt, M., Loskyll, M., & Zühlke, D. (2014). Human-machine-interaction in the Industry 4.0 era. In *12th International Conference on Industrial Informatics* (pp. 289–294). Porto Alegre, Brazil: IEEE Xplore.

Johnson-Laird, P. N. (2005). Mental models and thought. In K. J. Holyoak, & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 185–208). Cambridge, UK: Cambridge University Press.

Kablan, Z., & Erden, M. (2008). Instructional efficiency of integrated and separated text with animated presentations in computer-based science instruction. *Computers & Education*, *51*(2), 660–668.

Kay, M. (2006). Adaptive automation accelerates process development. *Bioprocess International Magazine*, *4*(4), 70–78.

Kumar, N., & Kumar, J. (2016). Measurement of efficiency of auditory vs visual communication in HMI: A cognitive load approach. In *Proceedings of the International Conference on Advances in Human–Machine Interaction* (pp. 1–8). Bangalore, India: IEEE. https://doi.org/10.1109/HMI.2016.7449168

Li, X., Li, D., Wan, J., Vasilakos, A. V., Lai, C. F., & Wang, S. (2017). A review of industrial wireless networks in the context of Industry 4.0. *Wireless Networks*, *23*(1), 23–41.

Lee, J. (2015). Smart factory systems. *Informatik-Spektrum*, *38*(3), 230–235.

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for Industry 4.0 based manufacturing systems. *Manufacturing Letters*, *3*, 18–23.

Lupiena, S., & Lepage, M. (2001). Stress, memory, and the hippocampus: Can't live with it, can't live without it. *Behaviour and Brain Research*, *127*(1-2), 137–158.

MacDougall, W. (2014). *Industry 4.0: Smart manufacturing for the future*. Berlin, Germany: GTAI.

Meixner, G., Petersen, N., & & Koessling, H. (2010). User interaction evolution in the SmartFactory^{KL}. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 211–220). Dundee, UK: British Computer Society.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.

Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): Literature review and practical application discussion. *International Journal of Production Research, 46*(13), 3517–3535.

Nakajima, S. (1988). *Introduction to TPM: Total productive maintenance*. Cambridge, MA, USA: Productivity Press.

Oechsner, R., Pfeffer, M., Pfitzner, L., Binder, H., Müller, E., & Vonderstrass, T. (2002). From overall equipment efficiency (OEE) to overall fab effectiveness (OFE). *Materials Science in Semiconductor Processing, 5*(4-5), 333–339.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Paas, F., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*(4), 737–743.

Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., De Amicis, R., & Vallarino, I. (2015). Visual computing as a key enabling technology for industrie 4.0 and industrial internet. *IEEE Computer Graphics and Applications*, *35*(2), 26–40.

Romero, D., Bernus, P., Noran, O., Stahre, J., & Fast-Berglund, A. (2016). The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In I. Nääs, O. Vendrametto, J. Mendes Reis, R. F. Gonçalves, M. T. Silva, G. von Cieminksi, & D. Kiritsis (Eds.), *Advances in Production Management Systems: Initiatives for a Sustainable World* (Vol. 488, pp 677–686). Cham, Switzerland: Springer.

Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two? *Mathematical and Computer Modelling*, *38*(3-4), 233–244.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging*, *6*(1), 118–127.

Scott, D., & Pisa, R. (1998). Can overall factory effectiveness prolong Mooer’s law? *Solid State Technology*, *41*(3), 75–81.

Sniderman, B., Mahto, M., & Cotteleer, M. J. (2016). *Industry 4.0 and manufacturing ecosystem.* London, UK: Deloitte University Press.

Streans, P. (2013). *The industrial revolution in world history.* New York, NY, USA: Westview Press.

Subramaniam, S., Husin, H. S., Yusop, Y., & Hamidon, A. H. (2009). Machine efficiency and man power utilization on production line. In the *8th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communication* (pp. 70–75). Cambridge, UK: ACM Digital Library.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Villani, V., Saattini, L., Czerniak, J. N, Mertens, A., Vogel-Heuser, B., & Fantuzzi, C. (2017). Towards modern inclusive factories: A methodology for the development of smart adaptive human-machine interfaces. In *International Conference on Emerging Technologies and Factory Automation* (pp. 1–7). Limassol, Cyprus: IEEE Xplore. https://doi.org/10.1109/ETFA.2017.8247634

Wan, J., Zhang, D., Zhao, S., Yang, L., & Lloret, J. (2014). Context-aware vehicular cyber-physical systems with cloud support: Architecture, challenges, and solutions. *IEEE Communications Magazine*, *52*(8), 106–113.

Wan, J., Zou, C., Ullah, S., Lai, C. F., Zhou, M., & Wang, X. (2013). Cloud-enabled wireless body area networks for pervasive healthcare. *IEEE Network*, *27*(5), 56–61.

Wang, S., Wan, J., Li, D., & Zhang, C. (2016). Implementing smart factory of industrie 4.0: An outlook. *International Journal of Distributed Sensor Networks*, *12*(1), 1–10. https://doi.org/10.1155/2016/3159805

Wittenberg C. (2015). Cause the trend Industry 4.0 in the automated industry to new requirements on user interfaces? In M. Kurosu (Ed.), Human-computer interaction: Users and contexts. *Lecture Notes in Computer Science* (Vol. 9171, pp. 238–245). Cham, Switzerland: Springer.

Wu, F. J., Kao, Y. F., & Tseng, Y. C. (2011). From wireless sensor networks towards cyber physical systems. *Pervasive and Mobile Computing, 7*(4), 397–413.

Zhang, D., Wan, J., Liu, Q., Guan, X., & Liang, X. (2012). A taxonomy of agent technologies for ubiquitous computing environments. *KSII Transactions on Internet & Information Systems, 6*(2), 547–565.

Zhong, R. Y., Dai, Q. Y., Qu, T., Hu, G. J., & Huang, G. Q. (2013). RFID-enabled real-time manufacturing execution system for mass-customization production. *Robotics and Computer-Integrated Manufacturing, 29*(2), 283–292.

Authors' Note

All correspondence should be addressed to
Dr. Naveen Kumar
WS 134, Department of Design,
Indian Institute of Technology Delhi,
Hauzkhas, New Delhi-110016, Delhi, India
naveenno122k@gmail.com

or

Dr. Jyoti Kumar
WS 134, Department of Design,
Indian Institute of Technology Delhi,
Hauzkhas, New Delhi-110016, Delhi, India
meetjyoti@gmail.com or jyoti@iddc.iitd.ac.in

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi

Volume 15(1), February 2019, 79–99

Female Sex Workers' Use of Mobile Phones in India: Lessons in Effective Engagement

Subadra Panchanadeswaran
Adelphi University School of Social Work
Garden City, NY
USA

Ardra Manasi Unnithan
Independent Research Consultant
New York, NY
USA

Shubha Chacko
Solidarity Foundation
Bengaluru
India

Michael Brazda
Center for Aging Studies
University of Maryland Baltimore County
Baltimore, MD
USA

Natalie Brooks Wilson
Adelphi University School of Social Work
Garden City, NY
USA

Santushi Kuruppu
Independent Not-for-Profit Advisor
New York, NY
USA

Abstract: *Despite the widespread impact of mobile phone technology, research on female sex workers' (FSWs) access and use of this technology is sparse. This study aimed to document FSWs' subjective perceptions of the benefits of mobile phone technology as well as examine the potential for its use by nongovernmental organizations (NGOs). Employing a phenomenological study built on in-depth qualitative interviews and focus group discussions with 67 FSWs and 18 staff and leaders from local NGOs in Mumbai and various locations of the Karnataka State in India between January and May 2015. Findings revealed complex trajectories for FSWs regarding accessing, learning about, and using mobile phone technology. FSW participants highlighted enhanced social connectedness with clients and their communities while NGO personnel underscored the newer ways of engagement with FSWs with mobile phone technology. Study findings point to both opportunities and challenges posed by FSWs' extensive use of mobile phone technology.*

Keywords: *sex work, mobile phone technology, India, prostitution, female sex workers (FSWs), technology-based interventions.*

DOI: https://doi.org/10.17011/ht/urn.201902201609

VERTAISARVIOITU KOLLEGIALT GRANSKAD PEER-REVIEWED www.tsv.fi/tunnus

INTRODUCTION

Globally, information and communication technologies (ICTs) have profoundly changed the manner in which people communicate with one another and seek and share information and ideas (Lipset, 2013). A recent report by the International Telecommunications Union (ITU, 2017) highlighted the significant digital divides between developed and developing countries. And, despite their worldwide reach, researchers in the past have underscored the inherent inequities that ICTs present, ranging from access, autonomy, and skill to social support and purpose (Sinpeng, 2015), that continue to foster enormous stratification and negate the initial popular perceptions of equalizing diverse populations (Mendonça, Crespo, & Simoes, 2015). In recent years, mobile phones have emerged as a powerful communication tool and there is widespread acknowledgment of the power of mobile phones to transform people's lives, especially in places where resources are scarce (Donner, 2008). Further, mobile phones have presented a powerful opportunity for advancement to millions of women in developing nations.

While ample literature abounds on the perceived benefits of mobile phone technology for marginalized women in the developing world, empirical evidence—specifically on female sex workers' (FSWs) experiences with mobile phone technology—is only gradually emerging. In India, tele-density among sex workers in urban settings has been estimated at 97% (Sambasivan, Weber, & Cutrell, 2011). Researchers in the recent past have consistently documented a trend of mobile phone-based client solicitation among FSWs in India and have consistently pointed to the benefits of moving away from soliciting in public venues, thereby enhancing desired anonymity as well as reduced presence of intermediaries such as brothel owners, brokers, and other middlemen (Beattie, Bradley, Vanta, Lowndes, & Alary, 2013; Buzdugan et al., 2010; Buzdugan et al., 2012; Jain & Saggurti, 2012). Given the existing evidence of widespread use of mobile phone technology by FSWs in India, it is vital to examine specifically the trajectories of sex workers' intersections with technology. Additionally, NGOs (nongovernmental organizations) and other agencies must understand and recognize sex workers' abilities to engage with this technology, as this will have implications for designing holistic interventions and outreach programs for women in high-risk and vulnerable groups. This study thus aims to fill this gap in literature by shedding light on how FSWs' experiences with accessing and using mobile phone technology may be similar or different compared to other women from marginalized communities. The focus of this qualitative study was two-fold: (a) to document FSWs' subjective perceptions of the benefits of mobile phone technology, and (b) to examine the potential for the use of mobile phone technology by NGOs.

To understand the specific ways in which FSWs interact with mobile phone technology, it is crucial to examine the existing terrain and nuances of how marginalized women from various social and cultural contexts employ mobile phones to their advantage. Research evidence has underscored the ways in which ICT technologies such as MP3-type players have enabled rural women's increased engagement in important community dialogues (Sengupta, Long, Singhal & Shefner-Rogers, 2007) and emphasized that independent access to mobile phone technology has enabled women's entrepreneurial endeavors while reducing the gender divide (Jacobsen, 2011; Jennings & Gagliardi, 2013; Zainudeen, Iqbal, & Samarajiva, 2010). In addition, for low-income women, finding intermediaries such as family members, including children and spouses, ably sidesteps some formidable barriers posed by literacy-related issues (Lemish & Cohen, 2005; Sambasivan, Cutrell, Toyama, & Nardi, 2010). Mobile phone technology also has been touted as a key enabler in the deepening of social ties among women, thus facilitating both interpersonal and instrumental gains through building of social capital. Research in the past has shown that women use mobile phones to initiate virtual relationships more frequently as compared to men (Pertierra, 2005). Several studies have acknowledged the widespread feminization of mobile phones and underscored women's use of the technology as being consistent with their caregiving roles. Some researchers have found that women use technology largely for intrinsic (relationship maintenance) rather than extrinsic purposes (business transactions; Lemish & Cohen, 2005; Zainudeen et al., 2010), while others noted that mobile phones symbolize safety and connection to family for women, enabling expressions of empathy (Livholts & Bryant, 2013). Research evidence also points to the role of mobile phone technology in strengthening of ties among household members, between members in a social circle, between employers and their employees, and more specifically between women employed overseas and their children (Handapangoda & Kumara, 2013; Murphy & Priebe, 2011; Pertierra, 2005; Tenhunen, 2008). Other studies have staunchly stressed mobile phone technology's indispensability for the expansion of women's social and business networks (Handapangoda & Kumara, 2013; Murphy & Priebe, 2011). In a recent global survey by GSMA (mWomen Programme, 2012a), the majority of women (80%) from impoverished communities reported greater familial connectedness because of mobile phones and highlighted the usefulness of mobile phones during emergencies. Other researchers have emphasized that these types of expanded access to others within communities serves to reduce significantly the women's sense of isolation (Maleka, 2012).

Findings from studies on the perceived impact of mobile phone technology on women's personal lives have been mixed. Some researchers have underscored the role of portable communication devices in fulfilling women's social needs in terms of extending their social networks and support structures, as well as fulfilling their entertainment needs by providing them access to movies and shows on topics of interest to them (Handapangoda & Kumara, 2013, Jouhki, 2013; Yoon, 2006). In Afghanistan, providing women access to culture-specific content related to civic education and development and rights-based frameworks via solar-powered digital audio players facilitated collective listening with family members, fueled vital discussions, defused potential male resistance, and promoted women's empowerment (Sengupta et al., 2007). Mobile phones also have been associated with increased safety and security for women, along with improved access to financial, health, and educational opportunities and women's increased autonomy and proactive decision making regarding their own and their families' health (GSMA Development Fund & Cherie Blair Foundation for Women, 2010).

One of the most important gains resulting from mobile phone technology has been in the public health sphere. Mobile phone technology has increasingly proved to be useful for increasing women's utilization of primary health care services (Oluwafemi & Wynn, 2014), implementing mental health interventions (Norris, Swartz & Tomlinson, 2013), and obtaining easy access to vital health-care information (Jennings & Gagliardi, 2013). Further, mobile phones have emerged as one of the most popular and low-cost technological devices used to disseminate health reminders and disease prevention messages in an unobtrusive manner, monitor treatment adherence, enable disease surveillance (including diagnosis and treatment of sexually transmitted infections), collect data regarding on-going health concerns, and provide immediate feedback to health-care practitioners (Bahadur & Murray, 2010; Patrick, Griswold, Raab, & Intille, 2008). Rapid advances in mobile phone technology are driving the increased prominence of mobile phones, allowing individuals to take more control over their health and helping health-care practitioners deliver timely and effective services. Innovations such as telephone broadcasting systems have aided in relaying psycho-educational information to women living with or at risk for HIV/AIDS, such as FSWs (Sambasivan, Weber & Cutrell, 2011), while inventive mHealth approaches have been proven to have positive health outcomes for maternal and child health (Deshmukh & Mechael, 2013).

On the other hand, researchers in the past have cautioned against over emphasizing the positive impacts of mHealth interventions on gender relations and have called for a comprehensive understanding of women's abilities to navigate sociocultural, financial, and technical literacy barriers in terms of access to and use of technology (Jennings & Gagliardi, 2013; Natarajan & Parikh, 2013). And, while the gains of mobile phone technology are undeniable, research evidence suggests a gender inequality owing to the gendered use of ICTs in general and mobile phones in particular (Casado & Lasén, 2014; Goluboff, 2016; Jouhki, 2013; Svensson & Wamala Larsson, 2016). Past studies have pointed to technology-related vulnerabilities for women often because of male-controlled access and monitoring of mobile phones or a lack of privacy resulting from phone sharing for cost effectiveness (GSMA mWomen Programme, 2012a; Pertierra, 2005), while others have drawn attention to the gender disparity regarding mobile phones with exclusive access and use by men for communication and entertainment purposes contrasting with women's experiences of monitored phone use by extended family members and with limited privacy (Doron, 2012). Many studies have found that women's overdependence on men and fear of reprisals from male spouses/partners, combined with a lack of training in using technology, were significant barriers to women's success in using technology effectively and expanding social networks outside of existing boundaries (GSMA mWomen Programme, 2012b, Handapangoda & Kumara, 2013; Lemish & Cohen, 2005; Ojokoh, 2009; Sambasivan et al., 2010). Archambault's (2011) study revealed the significant impact of the information gathered on and through mobile phones on young couples in Southern Mozambique that resulted in conflicts and relationship disruption. Other research suggested that the perceived anonymity of texting may sometimes inspire uncharacteristically intimate conversations, leading to premature or inappropriate self-disclosure or even to harassment (Pertierra, 2005; Short & McMurray, 2009), an issue that has specific implications for sex workers in particular, given the high premium placed on preserving their anonymity.

Significant research evidence illustrates both benefits and challenges in using mobile phone technology for those engaged in the sex work industry. Mobile phones and the Internet have been underscored as important new pathways for soliciting new clients and enabling ongoing client–sex worker interactions in Estonia (Aral, St. Lawrence, & Uuskula, 2006) and Thailand (Veena, 2007) and for establishing and maintaining transactional sex practices in sub-Saharan Africa (Stark, 2013). Research evidence from India has stressed the association between mobile phone use by FSWs for client solicitation and risky sexual behaviors (Mahapatra, Saggurti, Halli, & Jain, 2012; Navani-Vazirani et al., 2015), especially in the case of FSWs belonging to the lower socioeconomic strata who solicited clients both over the phone and at traditional venues (Navani-Vazirani et al., 2017). In a recent study in India, Panchanadeswaran, Unnithan, Chacko, Brazda, and Kuruppu (2017) emphasized FSWs' heightened vulnerabilities, including risk for conflict and violence within intimate relationships resulting from mobile phone surveillance by their partners. Support also exists for the idea that mobile phone technology has aided users' connections to others in addition to saving time, increased mobility, and in the dissemination of vital information from governmental organizations and NGOs (Jouhki, 2013; Millanga, 2014; Ohme, 2014). In terms of service delivery to the most socially marginalized populations like

FSWs, NGOs often find it hard to reach them, and they struggle with the existing means to track and monitor the progress of well-being among these groups through traditional forms of engagement. Research on phone-based systems and technology provide outreach actors such as NGOs with comprehensive strategies for improving the implementation of technology-mediated interventions, and mobile phone technology as a means of psycho-education is becoming increasingly common. Studies in this area have explored the impact of transmitting information to engage better with marginalized populations (Smales, 2011; Sambasivan et al, 2010; Sambasivan et al., 2011). Sambasivan et al. (2011) studied the use of a phone broadcasting system among FSWs to improve NGO outreach efforts in India to communicate information on issues ranging from microfinance loans and HIV testing to computer training and community-based information, while Smales (2011) discussed a text messaging hotline system implemented in Kuwait for domestic workers to follow-up on reports regarding gender-based violence. Overall, a convergence in the literature has developed regarding the central role that NGOs play in enabling communities with human–computer interactions, especially in the developing world (Sambasivan et al., 2010).

While literature on women's access and use of mobile phone technology in developing countries is abundant, focus on the experiences of FSWs' use of mobile phone technology and the implications for NGOs that work with this marginalized population is still emerging. Researchers have pointed to the need to examine mobile phone technology's impact beyond traditional areas of development, including health care, education, and so forth (Archambault, 2011; Sariola, 2009). The current study addresses this gap by delving into how the use of mobile technology among FSWs is unique, owing to the nature of sex work and where sex workers value a high degree of anonymity. The aim of the study was to examine the myriad ways in which mobile phone technology impacts the lives of FSWs in India. This paper will focus on two specific questions: (a) What are FSWs' subjective perceptions of the benefits of mobile phone technology? and (b) What role does mobile phone technology play in the ways in which FSWs and NGOs engage with each other?

METHODS

Design, Sampling, and Data Collection

For this study, we used a phenomenological approach to investigate the experiences of FSWs in Mumbai and various locations in Karnataka state, including the city of Bengaluru and several semi-urban and rural areas around Bengaluru. FSWs were recruited into the study by the first author in partnership with local NGOs and local professionally trained social workers. The eligibility criteria for the participants of the study were women who (a) were 18 years or older, (b) self-identified as sex workers, and (c) currently owned a mobile phone. The FSWs who participated in our study solicited and provided sexual services in multiple venues, primarily (a) street, lodges, or hotels, (b) their own or their friends' homes, or (c) brothels. A total of 67 FSWs were recruited into the study through purposive and snowball sampling methods. The mean age of the FSWs was 33.05 years (SD = 7.02 years); over a third (38%) had not attended school, and the majority (76%) reported being currently single with at least one child (M = 1.8, SD = 0.98). For most of the FSWs (85%), sex work was the primary source of income. The

FSWs reported owning mobiles for almost 9 years on average. A significant proportion (75%) of the FSWs owned one mobile phone, while 25% reported owning two mobiles. Eighteen individuals from local NGOs in Mumbai and Karnataka participated in our study, including front line staff (grassroots-level worker), supervisors, managers, and heads of organizations. Most (72.2%) were graduates in the fields of social work or related disciplines with experience in working with sex workers.

Data collection methods included in-depth interviews (44 FSWs) and several focus group discussions (23 FSWs), in addition to a short sociodemographic survey. The semistructured interview guides used for data collection were developed by the first author in consultation with local NGO partners and gatekeepers. The guides were piloted, revised, and subsequently finalized. Questions included initiation into mobile technology use, the range of mobile phone use, perceived benefits of mobile phone use on personal life and on sex work, and their experiences with using phones for engagement with NGOs as well as readiness to send/receive health promotion messages, other generic information, and so forth. In-depth interviews and focus group discussions were conducted by the first and the third author with the help of a research assistant (all female) in a private office of the participating NGOs. Interviews and discussions were conducted in a conversational style, generally after the interviewer and the respondent chatted informally over coffee/lunch for at least a half hour before commencing the data collection process. Data were collected in the local languages, that is, Hindi, Marathi, Kannada, or English (in which both the respondents and the researchers were fluent). NGO personnel provided their perspectives via in-depth interviews. All interviews and discussions were recorded after the interviewers administered detailed informed consent procedures. Based on local cultural norms, all FSW participants were compensated in kind (i.e., a meal, transportation costs to the data collection site, and a small gift). This study received approval from Adelphi University’s Institutional Review Board.

Data Analysis

Qualified multilingual transcribers translated and transcribed verbatim all non-English interviews into English for analysis and coding. Coding is the process of identifying themes in texts, then labeling the presence of these themes in the collected documents used for the research (Bernard, 2011). The authors developed a sample coding plan based on the recommendations of Miles and Huberman (1994). Code families containing primary and secondary codes were developed based on the goal of the study along with prior research input from NGO collaborators. The coding plan was used to team code four interviews as a pilot test. During this pilot process, the coded documents were discussed by the coders with specific attention paid to the code list and its ability to be used to document the target themes relating to the implications of mobile phone technology on sex workers lives (Miles & Huberman, 1994). The authors added, subtracted, and organized the code families and subcodes as necessary to clarify the data analysis process for use in the final coding of the transcripts (Miles & Huberman, 1994). The final step in the coding process was loading the transcripts into Atlas.ti 7 (Muhr, 2013) for analysis and coding.

Five code families were used in the analysis of the total dataset. Each family contained one to five primary codes and multiple secondary codes. In this paper, we focus on the analysis of the primary and secondary codes from two of the code families. These families are FSWs journeying into and benefits of mobile phone technology, and NGO engagement with sex workers using mobile phone technology. The primary and secondary codes in the journeying into technology family identified themes about how and why FSWs started using mobile phones, the costs associated with and the learning process to use mobile phone technology, and the perceived benefits of using the technology. The NGO engagement family contained themes related outreach activities using phones and implications for service delivery to the sex worker community.

FINDINGS

The main results from the study are presented below in two parts. The first part will focus on FSWs' pathways to accessing and learning about and use of mobile phone technology, as well as FSWs' views on the perceived benefits of mobile phone technology. The second part will highlight FSW respondents' and NGO personnel's opinions on the use of technology for optimal mutual engagement.

Journeying into Mobile Phone Technology

Initiation into Technology and Learning

Results revealed varied scenarios regarding how FSWs were introduced to mobile phone technology. In some cases, this initiation was sudden; in other instances, it entailed a gradual process of learning and familiarization. A 43-year old street-based FSW described how she has been using one or another form of mobile phone technology for the past 20 years:

> *It's been about 15 to 20 years since I first started using a mobile…. I first bought a pager, then a mobile with an antenna, then another one without a camera. And just recently, I bought this big mobile phone with a camera.*

For many respondents, mobile phone gifts from regular trusted partners and clients were a key gateway that facilitated access to the new technology and some of them recounted how they gained gradual comfort and familiarity using the same. As one 25-year old brothel-based FSW explained,

> *One of my customers gifted me. He said, "You use it. Slowly, slowly you will know everything." Then slowly, slowly it was easy for me to operate, and now I know everything. If I cannot contact over call, then I can contact through WhatsApp also.*

For others, watching peers gain from the access to mobile phones motivated them to proactively acquire one for themselves. Another brothel-based 30-year old FSW described her experience as follows,

> *I did not know the use of mobile phones initially. I used the landline and the coin booth for work. Later on, when I started seeing a lot of them (other FSWs) having it [mobile phone], I wanted one. But I couldn't buy one because I did not have money. When the rates came down, I bought.*

For a few FSWs who were employed with local NGOs on a part-time basis, these organizations became a key pathway for their induction into mobile phone technology. A 24-year old home-based FSW explained,

> *I didn't know anything about mobile when I started working with the NGO. They had given me a crisis phone, and when any crisis happened, they used to call me. I didn't know how to use it at first. I worked for nearly 7 years with that NGO. Then I gave up the crisis phone* [and] *bought one for myself. Meanwhile, I had learned everything.*

Many brothel-based FSWs also recounted instances in which the brothel owners urged them to buy phones to facilitate their ability to stay in touch with their families. A 31-year old FSW in a Mumbai brothel clarified,

> *No, I didn't purchase it immediately. When I was facing problem in contacting my children, then Mai* [brothel owner] *provided me a mobile phone with a SIM card. She said to me, "Keep this mobile phone now; later on I will provide you a Chinese Mobile and you repay me later on."*

Findings revealed that, over time, FSWs gained competencies in using mobile phone technology. For many, the phone seller was a key initiator and enabler who also provided technical support to help FSWs navigate the technology. Many FSWs described their experiences in ways similar to this respondent, who was a 35-year old brothel-based FSW:

> *If I go to the mobile shop, he shows, "This is how it opens; this is how you need to fix it." He* [phone seller] *gave me that piece* [headphones]. *But now looking at other women, I've learned to use a headphone in my ears, listen to music, and use WhatsApp. Everyone uses all this, and I learned* [by] *watching them use these. Earlier, I couldn't understand all this.*

A 36-year old female NGO manager asserted her observations of FSWs' learning from multiple actors,

> [They learn] *from each other. It is peer education. And some of these girls will learn from their customers. And there are many mobile shops around the red-light area. When they have some issues with the mobile phones, they will go and will check there and repair their phones and get some information from the shop owner as well.*

FSW respondents repeatedly identified their children and peers as key enablers of their learning to use mobile phone technology as did NGO personnel. A 48-year old female NGO manager stressed,

> *They are now also learning from their children. Children are studying, going to school.… Small children learn all these apps and they are teaching their mothers. ... They learn from each other and they learn very quickly. All these touchscreen phones, the android phones now, they operate very well….Even though they* [the FSWs] *are barely literate, they can identify names, numbers.…They have some kind of visual thing and they are able to do that.*[1]

For some FSWs, spouses and clients appeared to play a key role in helping them to get familiar with the various features of mobile phones, as evidenced by the following quotes:

> *If I have to write something or I want to talk to my friend or I want to send my friend something, I tell my husband to show what has to be done.* [38-year old street-based FSW]

> *Whatever knowledge he* [client] *had about mobiles, he taught me everything.* [42-year old street-based FSW]
>
> *Switching on/off Bluetooth, how to use WhatsApp, how people speak on WhatsApp, my client taught me all this. I still remember his name.* [smiles; 28-year old street-based FSW]

Perceived Benefits of Mobile Phone Technology

FSW participants reported that, regardless of the pathways through which they were initiated into mobile phone technology, mobile phones offered important uses in their lives, ranging from better connectedness with families, peers, and others in their communities to providing opportunities for their own enjoyment. Mobile phone technology was touted by many respondents as freeing them from the inconvenience of public phone booths. One home-based, 29-year old FSW narrated how she purchased mobile phones to escape from the tedium and disadvantages of using privately operated small phone shops (known as Subscriber Trunk Dialing or STD booths), especially at night:

> *I used to contact my customers or my family members through the coin phone. If I wanted to get in touch with someone in the middle of the night, based on my status* [as a sex worker], *it was difficult. Suppose I wanted to send someone a message in the night, I had to go looking for a coin phone booth or wait for the person to open his shop in the morning. Based on all these circumstances, I felt, why not own one myself? Hence I purchased it.*

Mobile phones appeared to be a vital facilitator in sex workers' efforts to rekindle and deepen ties with their loved ones. A 40-year old street-based and a 30-year old brothel-based FSW, respectively, recounted their experiences as thus:

> *A lot of changes have taken place. I am reunited with my family because of my mobile phone. I left my family about 9 years back. If I have good relationships with my family, then it is because of the mobile phone.*
>
> *Earlier, I didn't have much information about what's happening at home, so I would just stop thinking about it. There was no phone. Now, I get calls. Someone or the other calls up. Twice a week my brother calls, then my sister calls. So the relationship has become stronger.*

Many FSWs also talked about their decisions to gift mobile phones to their family members, leading to an elevation in their status within their families. Our findings revealed the extensive use of mobile phone technology to build social networks and relationships with peers, neighbors, community leaders, as well as to seek help during times of crises.

Significantly, the FSW respondents' narrations also revealed extensive use of mobile phones as tools for self-directed entertainment and personal enjoyment, such as for listening to music, watching movies, playing games, and for taking pictures/videos of friends and family. One 27-year old home-based FSW said, "*We play games, we listen to songs, we listen to FM* [radio], *and then we see photos. We speak over the phone, talk to our relatives, send messages, we read messages*." A 24-year old brothel-based FSW stressed, "*I get some songs saved for the kids. I pay about Rs.40* [US$0.59] *to get the songs downloaded. They watch movies too. I get them downloaded; they watch the movies and then they delete them.*" Although rare, a few FSWs also shared their experiences of shopping online with their phones, as one 35-year old street-based FSW described,

> *It* [mobile phone] *is useful now to buy jewelry: what sort of jewelry and* [from] *which store to buy. A friend who works with me will send me the link on it* [the phone]. *I can see it clearly. They show me the latest trends that I can choose from. They* [Web sites] *also suggest options that will be good. My son has selected two of those.*

For many sex workers in this study, mobile phone technology appeared to provide immense gateways that facilitated learning and other creative forms of self-expression. As one 30-year old brothel-based FSW asserted,

> *Mobile has changed my life. I may be in the wrong line* [of work], *but I've learned to read and write using a mobile. I've learned a lot through the mobile. Like, I didn't know how to sign. I didn't know what A looked like and didn't know any spellings. I didn't go to school at all. Slowly I learned this is A, B, C, D, etcetera*. [Writes her name.] *I've learned this because of the mobile.*

NGO personnel observed that sex workers' enhanced self-worth has resulted from mobile phone technology. This perspective is apparent in many of their narrations, as evidenced in the quotes below:

> *When someone has access to a phone, it is so empowering. … A few words of English is empowering;…they can negotiate.* [33-year old female front line worker]
>
> *Many can't read or write, so even the way they save a person's number through different ringtones, different visual symbols, etcetera., sort of makes them* [sex workers] *feel a sense of accomplishment and pride—that you don't have to be literate to be smart and intelligent.* [28-year old female supervisor]
>
> *Around 92-95% of the women in the red-light area are illiterate. But they know how to use the smart phone! Sending video clips, photographs through Bluetooth using Internet, and downloading songs & movies. Now they know to do that very easily.* [49-year old female NGO manager]

Another real advantage of mobile phone technology that FSWs spoke about at length was related to their experiences within the realm of sex work. Respondents repeatedly underscored the newer opportunities presented by mobile phone technology in soliciting customers. As one 29-year old male NGO manager said,

> *Especially with young girls, it's* [mobile phone technology] *a craze! They find it very necessary to have a smart phone. There are a few girls who have pornographic video clips on their cell phones. It becomes a very interesting way to solicit customers.*

Further, mobile phones appeared to provide FSWs a real chance to actively vet potential clients before engaging with them. A 45-year old street-based FSW elaborated on this process:

> *We call and find out who he is, and then ask him what color clothes is he wearing. On reaching the spot, we observe him for a long time and then give him a missed call and check if it is him. If we feel that the person is inappropriate then we don't go. That is what we do. We switch off the phone for a while.*

Our study data were replete with instances of FSWs using mobile phone technology also to stay safe, vigilant, and seek support to get out of potentially dangerous situations. Mobile phones were cited as critical for FSWs in building mutually supportive networks, as is evidenced from the following quotes from a 48-year old street and a 35-year old home-based FSW, respectively,

> *Once it so happened, a girl, who went to a lodge to serve a client, found there were four inside the room. She sent me blank messages. It was a code, I had told her that if there is an emergency and if she is unable to call, she should send me blank messages, and I will call her back or do something.*

> *We are a group of 4 to5 women....We usually inform at least one of them that I am going to this place with this person, so she will know. So, when I don't come back even after half an hour or one hour, she will start calling me. If I don't answer or if my mobile is switched off, she will start getting scared. Either they come searching for us or call the client and ask: "Why are you keeping her so long? Send her back." And they try to save us. We have done this many times.*

Despite the many advantages of mobile phones that FSWs cited, they also highlighted significant challenges. One of the key issues that FSW participants consistently pointed out was the cost of maintaining a mobile phone. Respondent narrations highlighted FSWs' elaborate strategies to keep the costs low. A 25-year old street and 20-year old brothel-based FSWs described their efforts as such:

> *I always keep a track on the costs per call and other offers. Based on the various mobile phone offers* [and] *opportunities, I choose the plan. There are offers for 3 months, 6 months, and 9 months. I know a person who owns a mobile recharge and SIM-providing shop. I go to him and ask him, which is the best recharge offer; if in case I need a SIM, I tell him that I need a good number. It should be an easy number for the others to keep note and for me to keep it by heart. If there is such a number, please keep it aside. I will purchase it. He calls me when there is good recharge plan or if there is such a SIM. I purchase it.*

> *Internet – when I feel I don't require it or it doesn't work,* [then] *I don't recharge. Even if I have recharge, I switch off* [the] *Internet on the phone.*

Another common strategy that all FSW respondents reported adopting was placing the onus of expensive talk-time on clients. A 41-year old street and a 30-year old home-based FSW strongly underscored their strategy:

> *Even if they recharge my phone, I give them a missed call only… I don't call them…They call back.*

> *Yes, customers normally call us only when required. So, if I give them a missed call, they could note down my number and call me when required.*

An aspect of mobile phone technology that posed significant hurdles for FSWs was the persistent fear of stealth photography and video recording by strangers and clients, and thus the resultant violation of their privacy. As a 34-year old brothel-based FSW explained,

> *There is a building nearby.... from where photos have been clicked and have been printed in the newspaper as well. If they hide in the building and click, nobody will get to know. This has happened a lot.*

These kinds of incidents pushed FSWs to stay vigilant at all times, especially during activities associated with their work, meaning during solicitation and encounters with clients. A brothel-based FSW, aged 28 years, shared her experience as follows:

> *One day an incident happened. A customer* [client] *was trying to click a video when I and my friend was* [sic] *talking. Later on, I asked that person, "What video have you clicked?" He said, "I have not clicked any video." Then I said, "Show me your mobile." He replied: "No, no, why should I show my mobile to you?" Then our manager checked his mobile and got some videos.*

For many FSWs, incessant interrogations and monitoring of their use of mobile phones by their intimate partners/spouses presented additional challenges. A 36-year old street-based FSW described her experiences with her husband:

> *Sometimes I think if we didn't have a phone, it would've been better.* [There would be] *no fights at home, no tension about the village. There would be nothing. Whenever I call* [people, they ask me to] *give money. Whenever I call, there's this problem [or] that problem. If someone calls* [me], *then there are questions* [from my husband]: "*Who called? Why?" Rather than answer these questions, we should not have a phone at all.*

NGO personnel who interacted with FSWs on a regular basis commented on the gradual rise in partner violence arising from sex workers' use of mobile phones. A 45-year old female NGO supervisor said,

> *Every day when they* [FSWs] *go back home, they* [partners] *will check the numbers on their phones. They will check how many calls, who has called. So it is a constant struggle for them* [FSWs] *to have those incoming and outgoing calls. They have to delete* [that information] *every day."*

A 51-year old female NGO manager pointed out additional disadvantages for brothel-based FSWs. The mobile phones allow for added surveillance and tracking of their activities and movements that, at times, results in a severe clamping down on their freedoms and privacy.

Mutual Engagement Using Mobile Phone Technology: Perspectives of FSWs and NGO Personnel

Our study findings highlighted the changes brought about by the advent of mobile phone technology with respect to the potential implication regarding engagement with the sex worker community, outreach, and service delivery. The NGO personnel who participated in our study touted the vital role played by the technology in their ability to respond to FSWs' calls for help in times of emergency in real time.

NGO respondents also underscored the ease of ongoing contact and outreach with the sex worker community because of extensive mobile phone usage by FSWs. A 44-year old female NGO supervisor noted,

> *It has become useful for us because keeping in touch is easier. Earlier, we had to go and hunt for them in brothels. Then we lost track of them. Now it has become easier: We have the mobile number. So, one call and you know where they are.*

In addition to these benefits, NGO personnel described the ease with which they could mobilize sex workers for protests and rallies through mobile phones, touting the technology's vital role as tools for grassroots activism. The same NGO manager continued,

> *The other thing* [mobile phones] *have helped is sex worker rights and activism. If there is a rape, to mobilize the community is infinitely easier…if you are doing a protest or a rally the way other people use it. Sex workers also have multiple identities. So it is not just work, not just leisure, and not just activism.*

Data from FSW participants underscored these points, with some strongly citing the role of mobile phone technology essential in reaching out to their peers and allies for events. One 43-year old street-based FSW claimed,

> *Of course it is useful! I can contact so many people over the phone for programs. I can mobilize people, both women sex workers and sexual minorities. My mobile phone is my best friend; it is like my full-time partner, helps me with everything. I can mobilize more than 100 people over the phone: My tongue has such a value and power; I have built a rapport in such a way with both the communities. I make sure that the people who are dependent on me, nothing bad should happen to them.*

Respondent narrations also shed light on mobile phone technology's ability to build bridges between NGOs, sex workers, and service providers, especially from the public sector, criminal justice system, and social services. A 31-year old male NGO head asserted,

> *We have a very strong outreach program. Recently, we had a health check-up camp. Three days before that, our staff went out to remind them* [FSWs]. *Several women told our outreach workers: "In this heat, why do you come to tell us? You have our cell phones. Why don't you just call us? Even on the day of the camp, the minute you call us we will come." So you see now how our outreach perhaps has become outreach through telephone.*

At the same time, some NGO leaders lamented how mobile phone technology had significantly altered the nature of engagement with the sex worker community, with fewer in-person visits and personal connections. A 42-year old female manager stated,

> *Over the phone, a lot of things can be sorted out. But in social work,* [in] *the one-to-one relationship that is required, the technology blocks* [that interpersonal connection]. *Seeing the facial expressions, hearing her* [FSW's] *struggles* [and] *troubles and what she is going* [through] *is important. Over the phone, I am only hearing her voice and am very technical in my approach. Yes, getting in touch with the woman* [FSW] *has become easier for the staff. For me, to an extent* [there is] *a little negative component because the field visits have become less* [frequent and the] *link to the community may break at times. But at other times, it makes the work much more easier because, with a limited number of field staff, that you can deal with more number of women* [FSWs].

A significant number of FSWs also considered mobile phone technology, especially phone calls, to be a more efficient and effective way for NGO personnel to communicate with them. This is well depicted in the following quote by a 32-year old street-based FSW, who described the newer interactions with local NGOs,

> *More than sending an SMS* [Short Messaging Service], *they* [the NGO staff] *give us a call and inform us about the program. With SMS, there is a possibility that people see it or don't see it. That is why they inform us over the phone. They explain the date, day, and venue, and give us the program details. They know that we don't see SMS. It is better to let us know over the phone.*

Many FSW respondents also acknowledged their literacy challenges and expressed their preference for voice calls. One 40-year old street-based FSW elaborated, *"Yes, voice message would be the best. One can just listen to it, one doesn't need to read. We know how to listen to voice messages."* Yet, despite obvious disadvantages, many FSWs expressed their strong affinity to learning technology. As one 28-year old brothel-based FSW said, *"If we have to learn something new, why wouldn't we learn it? It is required."*

Our findings revealed NGOs used mobile phone technology in varied ways to work with the sex worker community, ranging from initiating mobile technology-based health promotion interventions, and crisis messaging to information dissemination on important events. A few NGO staff underscored the importance of working with brothel owners actively to support the FSWs during times of crises. A number of NGO respondents underscored the vital challenge in enabling FSWs' independence during times of emergency through mobile apps because of linguistic barriers. As one 34-year old male supervisor pointed out, the apps useful for safety *"are all in English, so it is very difficult* [for FSWs] *to write* [and] *type. So it is not easy to make them* [FSWs] *use all these apps…because they* [are] *not available in any of the regional languages."*

Importantly, a few NGO personnel respondents also reflected on the slow uptake of mobile phone technology among the NGO community, highlighting a distinct divide in the sex workers' and NGO staff's readiness to use and adopt technology for interventions. Another 45-year old NGO leader expressed candidly,

> *We are not advanced in technology, so we can't expect the women to be. We need to develop our own knowledge and get equipment* [phones] *to be able to deal with this type of situation. The older generation* [of staff] *tend to have technology phobia.*

Another significant aspect that especially NGO managers highlighted was the real gap between technology used by some FSWs and by NGO staff. This issue was highlighted by a senior female NGO manager, who said,

> *Ours is an absolutely local grassroots organization, so we have also staff from lower middle-class families…So, in terms of technology—their phones—we are limited…. The organization has to make arrangements for high-end phones if need to have it for the staff. You can't do it for one person; you need to do it for the entire* [staff]. *So, that poses a lot of difficulties in the process.*

Finally, a few of the NGO personnel expressed reservations about omnipresent availability necessitated by mobile phone technology. Their concern focused on the expectation that they would be available to respond to FSWs at all hours during the day and night.

DISCUSSION

Findings from the current study revealed the ubiquitous presence and salience of mobile phone technology in the lives of FSWs in India. Results revealed how FSWs in India appeared to have acquired, adopted and adapted the mobile phone to suit their own personal, social, and professional practices, as has been found in earlier research among low-income populations (Lipset, 2013; Tenhunen, 2008). As in the previous literature that pointed to low-income women with low levels of literacy finding intermediaries such as family members, including children and spouses, to help them to master mobile phone technology (Lemish & Cohen, 2005; Sambasivan et al., 2010), this study also shed light on some of the key actors responsible for the initiation of mobile phone technology among sex workers. These actors include their trusted partners/clients, peers, children, husbands, NGOs, phone sellers, and brothel owners.

Results highlighted some of the perceived benefits that FSWs attributed to mobile phone technology and the manner in which this marginalized group deftly adopted and adapted the technology to their advantage. The process of gradual familiarization with the new technology among FSWs revealed the role of mobile phone technology in enabling FSWs' sense of self-efficacy through mastering the use of mobile phones for self-directed enjoyment and skill building, including clicking pictures, consuming music, watching movies, shopping online, and learning English. Moreover, the FSWs' sense of deep appreciation for the device is similar to findings in earlier studies in India among nonFSW populations (Doron, 2012; Jouhki, 2013; Mehta & Mehta, 2014) and echoed Sengupta et al.'s (2007) study that showed the benefits obtained by poor women from owning electronic devices, such as MP3 players. FSWs' experiences in this study mirrored the experiences of lower-class women who worked as domestic workers—that is, their ability to connect with others for their businesses and so forth—but sharply contrasted with upper-class married women's supervised and limited use of the same within traditional households, as found in earlier research in India (Doron, 2012; Jouhki, 2013).

Most FSW respondents in our study were primarily self-employed and were able to exercise significant agency and choice in initiating ownership of new technology, thus maintaining control and proactively using their mobile phones to further their personal goals to a large extent. Further, the results from the narratives of FSWs in our study indicated a sense of pride and achievement owing to their adoption and use of mobile phone technology with ease, despite the lack of formal education, which is in line with the overall findings from past literature that have underscored exposure to technology and women's confidence (GSMA mWomen Programme, 2010; Lemish & Cohen, 2005; Sambasivan et al., 2010; Sengupta et al., 2007). Researchers have demonstrated technology's role in furthering women's access to productive employment and the resulting financial well-being (Boateng, Hinson, Galadima & Olumide, 2014; GSMA mWomen Programme, 2012a; Jennings & Gagliardi, 2013). Our results echoed those from past research, demonstrating that the poor FSWs' novel ways of using mobile phone technology to solicit clients aided in their efforts to achieve financial stability. Past studies have focused on women's use of mobile phone technology to expand their social networks and to maintain their connections with family members (GSMA mWomen Programme, 2010; Jouhki, 2013; Livholts & Bryant, 2013; Pertierra, 2005; Sengupta et al., 2007; Tenhunen, 2008; Zainudeen et al., 2010), which was a recurrent theme in the our study as well. Our FSW respondents used mobile phones to forge deeper and more meaningful ties within peer networks, community members, and their families, and were especially useful in seeking help during crises.

Despite the real gains of mobile technology for FSWs, findings from our study underscore the need for cautious optimism regarding the perceived impact of mobile phone technology for FSWs. Mobile phones appear to both liberate and constrain FSWs by fostering independence but yet somewhat isolate and perpetuate increased control of the FSWs by diverse actors in their lives. This study's findings clearly indicate mobile phone technology's capacity to challenge sex workers' need for privacy and anonymity because of nonconsensual photography and videography by clients and strangers. This finding was similar to earlier research in India among nonsex worker populations, which found that women were concerned about the "abuse of photography by men," referring to men secretly clicking pictures of women (Jouhki, 2013, p. 47). Although FSWs in our study appeared to exercise agency and autonomy in procuring and using mobile phones, their ability to feel empowered was significantly diminished within gender-bound intimate relationships, where surveillance stemming from their partner's sexual jealousy, and culture-specific gender norms in India was commonplace, as found in prior studies (Archambault, 2011, Doron, 2012; Jouhki, 2013; Panchanadeswaran et al., 2017).

Another striking finding from our study was the NGO personnel's recognition of the possibilities of creating technology-based interventions for the sex worker community, given FSWs' extensive adoption and adaptations of mobile phone technology, while acknowledging their own limited current use of the technology. Given that the technologies underlying mobile phones are becoming more powerful, cheaper, and increasingly accessible to all segments of populations globally (Patrick et al., 2008), mobile phones present an important opportunity to develop mobile phone-based mechanisms for greater outreach and service delivery to vulnerable groups such as sex workers. As prior studies show, the lack of local and community-related content can be a major barrier in women's use of IT for economic empowerment (Hafkin & Taggard, 2001; Jorge, 2002; Ojokoh, 2009).

CONCLUSIONS

A few limitations of the study bear mentioning. Given that all the data were based on self-reports, the disadvantages posed by recall bias and social desirability are real. Further, findings are based on an examination of the experiences and views of a group of purposively sampled FSWs and NGO personnel. As a result, it is not possible to generalize the conclusions drawn from this study to the larger population of FSWs in India. However, these limitations notwithstanding, this study fills an important gap in existing literature by highlighting the complex socio-cultural-economic considerations that govern the optimal level of FSWs' access to, adaptation of, and use of mobile phone technology.

This study adds to the existing literature in understanding the unique and nuanced ways in which FSWs acquire, adopt, and adapt mobile phone technology and how this is different from and similar to the use of mobile phones by marginalized women from distinct sociocultural contexts around the world. Thus, evidence from this study points to some important implications.

First, FSWs' extensive use of and ease with mobile phone technology in India presents an important opportunity for the development of phone-based interventions that address not only crises interventions but also help strategize and deepen social connectedness beyond their immediate communities. Mobile phones also may be a vital tool in helping sex workers to organize, mobilize, and advocate for change, as well as participate actively in local and regional political forums. Seamlessly integrating mobile phone-based tools into existing services for sex workers may be an urgent priority for NGOs (Navani-Vazirani et al., 2017). Most importantly, as underscored in prior research (e.g., Thomas et al., 2017), mobile phone-based interventions with FSWs need to outgrow merely addressing issues around HIV prevention and encompass domains such as digital literacy and building competencies around safety and privacy. As this phenomenological exploration suggests, while mobile phone technology has the potential to be a transformative social change agent in terms of providing FSWs with a sense of agency, it may not be necessarily empowering given its abilities to deepen gender divides and hierarchies. Thus, future research will need to focus on the manner in which ICTs, specifically mobile phone technology, impact gender relations for FSWs within various sexual partnerships.

More research that clearly delineates the barriers to optimal access and use of ICTs by women in general and FSWs in particular is critical in addition to recognition of women's agency in overcoming constraints (Chib & Chen, 2011). Designers of technologies for poor women in developing nations like India will need to account for the challenges faced by marginalized and stigmatized groups like FSWs around issues of privacy, literacy, and cost. Importantly, listening to women's voices and acknowledging their ideas in the quest to make ICTs accessible and ensuring their equitable use would be an essential prerequisite in developing newer interventions (Ojokoh, 2009) while remaining cognizant of the fact that significant social changes cannot be directly attributed to any technology (Mehta & Mehta, 2014). Further, understanding FSWs' intersections with mobile phone technology in the context of the similarities and the differences with other at-risk vulnerable populations is critical when developing and using technology-based interventions to affect change. It is critical to bear in mind that FSWs' access and use of mobile phone technologies are besieged with complex sociocultural and economic considerations. Given that FSWs greatly value privacy, trust, and anonymity, it is critical that mobile phone-based interventions recognize the inherent challenges to potential violation of privacy and respect sex workers' need for strict boundaries (Sambasivan et al., 2011) while balancing the need for personal connection. A thorough evaluation of the challenges posed by mobile phone technology, including linguistic and literacy barriers in applications that prevent the optimal use of technology by sex workers, would be critical in the development of mobile phone technology-based interventions.

IMPLICATIONS FOR APPLICATION

Technology is a multifaceted tool and has significant implications for sex workers' personal and professional lives. In addition to enabling increased agency for furthering their financial stability, mobile phone technology also plays a critical role in cementing FSWs' social ties with their family, friends, professional networks, and the larger community. Exploiting the ubiquity of mobile phone technology among the FSW community should be an important priority for policy-makers, NGOs, and other stakeholders for timely outreach, crisis intervention, and service delivery. Demystifying technology through ongoing training and updating technical skills would also go a long way in empowering marginalized FSWs to ensure that women gain control technology to maximize the benefits and reduce vulnerabilities inherent in the use of mobile phone technology.

ENDNOTE

1. In all the quotes, the use of ellipses indicates pauses in the speech of the participants.

REFERENCES

Aral, S. O., St Lawrence, J. S., & Uuskula, A. (2006). Sex work in Tallinn, Estonia: The sociospatial penetration of sex work into society. *Sexually Transmitted Infections, 82,* 348–353. https://doi.org/10.1136/sti.2006.020677

Archambault, J. S. (2011). Breaking up "because of the phone" and the transformative potential of information in Southern Mozambique. *New Media & Society*, *13*(3), 444–456.

Bahadur, M. K. C., & Murray, P. J. (2010). Cell phone short messaging service (SMS) for HIV/AIDS in South Africa: A literature review. *Studies in Health Technology and Informatics*, *160*, 530–534. https://doi.org/10.3233/978-1-60750-588-4-530

Beattie, T. S. H., Bradley, J. E., Vanta, U. D., Lowndes, C. M., & Alary, M. (2013). Vulnerability reassessed: The changing face of sex work in Guntur district, Andhra Pradesh. *AIDS Care, 25*, 378–384. https://doi.org/10.1080/09540121.2012.701726

Bernard, H. R. (2011). Qualitative data analysis I: Text analysis. In N. E. Roberts & M. Manzano (Eds.), *Research methods in anthropology: Qualitative and quantitative approaches* (5th ed.). Lanham, MD, USA: Rowman & Littlefield.

Boateng, R., Hinson, R., Galadima, R., & Olumide, L. (2014). Preliminary insights into the influence of mobile phones in micro-trading activities of market women in Nigeria. *Information Development*, *30*(1), 32–50. https://doi.org/10.1177/0266666912473765

Buzdugan, R., Copas, A., Moses, S., Blanchard, J., Isac, S., Ramesh, B. M., Washington, R., Halli, S. S., & Cowan, F. M. (2010). Devising a female sex work typology using data from Karnataka, India. *International Journal of Epidemiology*, *39*(2), 439–448.

Buzdugan, R., Halli, S. S., Hiremath, J. M., Jayanna, K., Raghavendra, T., Moses, S., Blanchard, J., Scambler, G., & Cowan, F. (2012). The female sex work industry in a district of India in the context of HIV prevention. *AIDS Research and Treatment*, Article ID 371482, 1–10. http://dx.doi.org/10.1155/2012/371482

Casado, E., & Lasén, A. (2014). What is disturbing and why not to disturb: On mobile phones, gender, and privacy within heterosexual intimacy. *Mobile Media & Communication*, *2*(3), 249–264.

Chib, A., & Chen, V. H. H. (2011). Midwives with mobiles: A dialectical perspective on gender arising from technology introduction in rural Indonesia. *New Media & Society*, *13*(3), 486–450.

Deshmukh, M., & Mechael, P. (2013). *Addressing gender and women's empowerment in mHealth for MNCH: An analytical framework*. Washington, DC, USA: mHealth Alliance. Retrieved on June 2, 2017, from http://www.villagereach.org/wp-content/uploads/2013/07/gender_analytical_framework_report.pdf

Donner, J. (2008). Research approaches to mobile use in the developing world: A review of the literature. *The Information Society: An International Journal*, *24*(3), 140–159.

Doron, A. (2012). Mobile persons: Cell phones, gender and the self in North India. *The Asia Pacific Journal of Anthropology*, *13*(5), 414–433.

Goluboff, S. L. (2016). Text to sex: The impact of cell phones on hooking up and sexuality on campus. *Mobile Media & Communication*, *4*(1), 102–120.

GSMA Development Fund & Cherie Blair Foundation for Women (2010). *Women & mobile: A global opportunity—A study on the mobile phone gender gap in low and middle-income countries.* Retrieved August 22, 2017, from http://www.cherieblairfoundation.org/wp-content/uploads/2012/07/women_and_mobile_a_global_opportunity.pdf

GSMA mWomen Programme (2012a). *Realising the mWomen Opportunity: A framework for designing the mWomen business case.* Retrieved September 1, 2017, from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2013/01/GSMA_mWomen_Business_Case_Framework_Full_Report_-FINAL.pdf

GSMA mWomen Programme (2012b). *Striving and surviving: Exploring the lives of women at the base of the pyramid*. Retrieved on September 1, 2017, from https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2013/01/GSMA_mWomen_Business_Case_Framework_Full_Report_-FINAL.pdf

Hafkin, N., & Taggart, N. (2001, June). Gender, information technology, and developing countries: An analytic study. Retrieved on September 22, 2017, from http://www.mujeresenred.net/zonaTIC/IMG/pdf/Gender_Book_NoPhotos.pdf

Handapangoda, W. S., & Kumara, A. S. (2013). The world at her fingertips? Examining the empowerment potential of mobile phones among poor housewives in Sri Lanka. *Gender, Technology and Development, 17*(3), 361–385. https://doi.org/10.1177/0971852413498742

Smales, P. (2011). The power to organise and engage: The use of ICT by women migrant domestic workers' organisations—Hong Kong and Taiwan (Research Program Report 2011). Retrieved on August 23, 2017, from https://www.gender-is-citizenship.net/sites/default/files/citigen/uploads/DraftTT_cover.pdf

International Telecommunications Union (ITU), Telecommunication Development Bureau. (2017). ICT facts and figures 2017. Retrieved on September 30, 2017, from https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf

Jacobsen, J. P. (2011). The role of technological change in increasing gender equity with a focus on information and communications technology. *ACSPL Working Paper Series*, *1, Article 2.* Available at https://wesscholar.wesleyan.edu/wps/vol1/iss1/2

Jain, A. K., & Saggurti, N. (2012). The extent and nature of fluidity in typologies of female sex work in southern India: Implications for HIV prevention programs. *Journal of HIV/AIDS & Social Services*, *11*, 169–191. https://doi.org/10.1080/15381501.2012.678136

Jennings, L., & Gagliardi, L. (2013). Influence of mHealth interventions on gender relations in developing countries: A systematic literature review. *International Journal for Equity in Health*, *12*(1), 85–95. https://doi.org/10.1186/1475-9276-12-85

Jorge, S. N. (November, 2002). *The economics of ICT: Challenges and practical strategies of ICT use for women's economic empowerment.* Paper presented at the United Nations Division for the Advancement and Empowerment of Women (DAW) Expert Group Meeting on Information and Communication Technologies and their Impact on and Use as an Instrument for the Advancement and Empowerment of Women. Seoul, Republic of Korea.

Jouhki, J. (2013). A phone of one's own? Social value, cultural meaning and gendered use of the mobile phone in South India. *Journal of the Finnish Anthropological Society, 38*(1), 37–58.

Lemish, D., & Cohen, A. A. (2005). On the gendered nature of mobile phone culture in Israel. *Sex Roles*, *52*(7-8), 511–521. https://doi.org/10.1007/s11199-005-3717-7

Lipset, D. (2013). Mobail: Moral ambivalence and the domestication of mobile telephones in Peri-Urban Papua New Guinea. *Culture, Theory and Critique*, *54*(3), 335–354. https://doi.org/10.1080/14735784.2013.826501

Livholts, M., & Bryant, L. (2013). Gender and the telephone: Voice and emotions shaping and gendering space. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments, 9*(2), 157–170.

Mahapatra, B., Saggurti, N., Halli, S. S., & Jain, A. K. (2012). HIV risk behaviors among female sex workers: Using cell phone for client solicitation in India. *Journal of AIDS and Clinical Research. S1:014*, 1–6 https://doi.org/10.4172/2155-6113.S1-014

Maleka, M. B. M. (2012). *A gender-based analysis of ICT adoption and usage in South Africa.* Master's Thesis, University of Witwatersrand, Johanneshburg, South Africa. Retrieved October 2, 2017 from https://core.ac.uk/download/pdf/39670837.pdf

Mehta, B. S., & Mehta, N. (2014). ICT and socio-economic empowerment of rural women: Case of mobile phone in India. *Knowledge Horizons–Economics*, *6*(4), 103–112.

Mendonça, S., Crespo, N., & Simões, N. (2015). Inequality in the network society: An integrated approach to ICT access, basic skills, and complex capabilities. *Telecommunications Policy*, *39*(3-4), 192–207.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA, USA: Sage Publications.

Millanga, A. K. (2014). Mobile phones and participatory communication for poverty eradication on public service broadcasting: The case of Tanzania Broadcasting Corporation (TBC). *Mobile Media & Communication*, *2*(3), 281–297.

Muhr, T. (2013). *ATLAS.ti 7.0 user's manual and reference* (Vol. 7). Berlin, Germany: Scientific Software Development GmbH.

Murphy, L. L., & Priebe, A. E. (2011). "My co-wife can borrow my mobile phone!" Gendered geographies of cell phone usage and significance for rural Kenyans. *Gender, Technology and Development*, *15*(1), 1–23. https://doi.org/ 10.1177/097185241101500101

Natarajan, M., & Parikh, T. (2013, December). Understanding barriers to information access and disclosure for HIV+ women. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development: Full Papers* (Volume 1; pp. 143–152). https://doi.org/10.1145/2516604.2516627

Navani-Vazirani, S., Heylen, E., Deardorff, J., Srikrishnan, A. K., Vasuedevan, C. K., Solomon, D., & Ekstrand, M. L. (2017). The role of sex work pay in moderating the effect of mobile phone solicitation on condom practices: An analysis of female sex workers in India. *HSOA Journal of AIDS Clinical Research & STDs*, *4*(1), 1–8. https://doi.org/10.24966/ACRS-7370/100008

Navani-Vazirani, S., Solomon, D., Gopalakrishnan, Heylen, E., Srikrishnan, A. K., Vasudevan, C. K., & Ekstrand, M. L. (2015). Mobile phones and sex work in South India: The emerging role of mobile phones in condom use by female sex workers in two Indian states. *Culture, Health & Sexuality*, *17*(2), 252–265.

Norris, L., Swartz, L., & Tomlinson, M. (2013). Mobile phone technology for improved mental health care in South Africa: Possibilities and challenges. *South African Journal of Psychology*, *43*(3), 379–388. https://doi.org/10.1177/0081246313493376

Ohme, J. (2014). The acceptance of mobile government from a citizens' perspective: Identifying perceived risks and perceived benefits. *Mobile Media & Communication*, *2*(3), 298–317.

Ojokoh, B. A. (2009). Empowering Nigerian women using information technology: Practical strategies. *EAF Journal*, *20*(1), 58–78.

Oluwafemi, S. O., & Wynn, R. (2014). Giving cell phones to pregnant women and improving services may increase primary health facility utilization: A case–control study of a Nigerian project. *Reproductive Health*, *11*(1), 8–16.

Patrick, K., Griswold, W. G., Raab, F., & Intille, S. S. (2008). Health and the mobile phone. *American Journal of Preventive Medicine*, *35*(2), 177–181. https://doi.org/10.1016/j.amepre.2008.05.001

Panchanadeswaran, S., Unnithan, A. M., Chacko, S., Brazda, M., & Kuruppu, S. (2017). What's technology got to do with it? Exploring the impact of mobile phones on female sex workers' lives and livelihood in India. *Gender, Technology and Development*, *21(*1–2), 152–167. https://doi.org/10.1080/09718524.2017.1385318

Pertierra, R. (2005). Mobile phones, identity and discursive intimacy. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *1*(1), 23–44.

Sambasivan, N., Cutrell, E., Toyama, K., & Nardi, B. (2010). Intermediated technology use in developing communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2583–2592). Retrieved on August 31, 2017, from https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/p2583-sambasivan-intermediated.pdf

Sambasivan, N., Weber, J., & Cutrell, E. (2011). Designing a phone broadcasting system for urban sex workers in India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 267–276). Retrieved on July 1, 2016, from https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/CHI201120Sambasivan20et20al.pdf

Sariola, S. (2009). *Gender and sexuality in India: Selling sex in Chennai*. Abingdon, UK: Routledge.

Sengupta, A., Long, E. G., Singhal, A., & Shefner-Rogers, C. L. (2007). The Sada says “We women have our rights”: A gender analysis of an ICT initiative in Afghanistan. *International Communication Gazette*, *69*(4), 335–353. https://doi.org/10.1177/1748048507079006

Short, E., & McMurray, I. (2009). Mobile phone harassment: An exploration of student's perceptions of intrusive texting behavior. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *5*(2), 163–180.

Sinpeng, A. (2015, October 21). More technology doesn’t mean less inequality [Web log post]. Retrieved on July 18, 2017, from https://theconversation.com/more-technology-doesnt-mean-less-inequality-48180

Stark, L. (2013). Transactional sex and mobile phones in a Tanzanian slum. *Suomen Antropologi: Journal of the Finnish Anthropological Society*, *38*(1), 12–36.

Svensson, J., & Wamala Larsson, C. (2016). Situated empowerment: Mobile phones practices among market women in Kampala. *Mobile Media & Communication*, *4*(2), 205–220.

Tenhunen, S. (2008). Mobile technology in the village: ICTs, culture, and social logistics in India. *Journal of the Royal Anthropological Institute*, *14*(3), 515–534.

Thomas, B., Closson, E. F., Biello, K., Menon, S., Navakodi, P., Dhanalakshmi, A., Mayer, K. H., Safren, S. A., & Mimiaga, M. J. (2017). Development and open pilot trial of an HIV-prevention intervention integrating mobile-phone technology for male sex workers in Chennai, India. *Archives of Sexual Behavior, 46*(4), 1035–1046.

Veena, N. (2007). Revisiting the prostitution debate in the technology age: Women who use the Internet for sex work in Bangkok. *Gender, Technology and Development*, *11*, 97–107. https://doi.org/10.1177/097185240601100105

Yoon, K. (2006). The making of neo-Confucian cyberkids: Representations of young mobile phone users in South Korea. *New Media & Society, 8*(5), 753–771.

Zainudeen, A., Iqbal, T., & Samarajiva, R. (2010). Who’s got the phone? Gender and the use of the telephone at the bottom of the pyramid. *New Media & Society, 12*(4), 549–566. https://doi.org/10.1177/1461444809346721

Authors’ Note

This research was funded in part by a Fulbright-Nehru Award for Academic and Professional Excellence given to the first author.

We thank all the participants for generously sharing their insights with us. We also thank the staff and leadership at Navjeevan, Mumbai; Aneka, Bengaluru; and the Karnataka Sex Workers’ Union; and all the participant NGO personnel for their participation in, commitment to, and support of the study. Special thanks to Prabhavati for her help with the study. We also thank Ms. Joanna Barberii and Ms. Natalie Brooks-Wilson for their help with the paper.

The authors have no conflicts of interest to disclose.

All correspondence should be addressed to
Subadra Panchanadeswaran
Adelphi University School of Social Work
1 South Avenue
Garden City, NY 11530, USA
panchanadeswaran@adelphi.edu

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

Human Technology

ISSN: 1795-6889

www.humantechnology.jyu.fi

Volume 15(1), February 2019, 100–135

USE WITHOUT TRAINING: A CASE STUDY OF EVIDENCE-BASED SOFTWARE DESIGN FOR INTUITIVE USE

Patrick Lehane
Information Services Division
University of Southern Queensland
Australia

Abstract: *This paper reviews intuitive software design and outlines the development of an instrument for analysts to evaluate the intuitiveness of software design. Current intuition research outlines three requirements for intuitive use: (a) existing experiential domain knowledge and skills, (b) an unexplainable perception that a novel situation is contextually familiar, and (c) successful application of users' previously acquired experiential knowledge and skills. A case study illustrates how these requirements can be specified, implemented, and evaluated. Questions to evaluate the characteristics of intuitive design and use resulted in an intuitive use evaluation of 3.2 on a scale of 0–4, indicating a perception of intuitive use. Subsequent factor analysis exposed three factors describing intuitive use: (a) Familiar User Expectations, (b) Confident Interactions, and (c) Leverage of Prior Learning. These factors map one-on-one to the requirements for intuitive use: providing an early confirmation of the proposed structure for analysis of intuitive software design and use.*

Keywords: *intuition, intuitive design, interaction design, usability, affordances, user experience, system analysis.*

DOI: https://doi.org/10.17011/ht/urn.201902201610

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

BACKGROUND

In this paper, I discuss the use of current research findings on intuition in an action research project involving software design and development. The criteria for intuition were used to develop a short survey for analysts to assess the degree of intuitiveness of a system design. The survey results can be included in stakeholder reports. The case study in this paper was the development of a course-level student evaluation (CLSE) survey system to be used by all divisions at a regional Australian university. Such a system used across the organization is termed an enterprise system. In this case, the enterprise system requires a manager to commission and manage the system on an ongoing basis and a cohort of academic staff users who use the system to promulgate content for the secondary cohort of student users. Regional universities in Australian have their principal campuses located outside the metropolitan areas of the states' capital cities.

The enterprise system used to deliver the CLSE has functionality that facilitates student feedback to university academic staff on the courses they run. The system's functionality also provides a feedback loop for academic staff to give feedback to the students on any comments made during the survey.

The software design brief from the university's senior management included the requirement that the software must allow users to complete the required tasks without formal training, a requirement more commonly specified as intuitive use. In this paper, I discuss the tasks required for a CLSE that would be employed by the academic staff. The system brief specified a CLSE that could be accessed by students on devices ranging from a desktop PC to an iPad, a survey management system for academic staff that could be used on similar devices, and a survey design and life-cycle management system for the survey system manager. Rather than build parallel systems for specific platforms, a responsive design system approach was adopted.

The challenge was to define intuitive use of the screen artifacts of a natural user interface for a responsive system. In the following chapter, I discuss the elements of intuitive use and current research on intuitive use. I then focus the discussion on how those tenets were applied to the design of an enterprise system used by academic staff who predominately operate desktop computers or laptops with keyboard and mouse in their offices. In this context, the visibility of interaction options is a paramount requirement for useful and consistent operations. Thus, a responsive design facilitated natural user interface operations with a touch screen. However, the commands to control the shaping of the data and progression through the business processes are menu driven.

Consequently, integral to determining the parameters for intuitive use was the need to specify the requirements for intuitive design and provide an instrument to evaluate whether or not intuitive use had actually occurred. This research paper presents a case study on intuitive design for intuitive use and the development of an evaluation method to determine whether intuitive interactions occurred.

THE USER INTERFACE

The typical consumer user interface nowadays is commonly called the graphic user interface (GUI). It replaced the black-screen command-line interface of the mid-to-late 20th century. Shneiderman (1983) brought attention to the interaction requirement of the GUI: The graphic elements were not autonomous and required an operator. Shneiderman coined the term "direct manipulation" to describe this operational requirement:

> The central ideas [of direct manipulation] seemed to be visibility of the object of interest; rapid, reversible, incremental actions; and replacement of complex command language syntax by direct manipulation of the object of interest—hence the term "direct manipulation." (Shneiderman, 1983, p. 57)

Because of this need for direct manipulation of GUI screen artifacts, good usability became a primary interaction requirement. Employing a direct manipulation interface (DMI) in screen design assists in focusing the design on this important aspect. It focuses the design on the interaction required for direct manipulation of the screen artifacts. Thus, the user interface is both a GUI and a DMI. The GUI facet is the vehicle to present to the user the screen artifact affordances, that is, their visual appearance and location. The DMI facet provides the underlying means to operate the commands and functionality of the software (e.g., a keyboard, mouse, touch pad).

Research into intuitive design and interaction is directed by both the visual aspects of the GUI design and the operational characteristics of the DMI design. The mental effort expended in working memory to use a computer is termed the cognitive load. The cognitive load comprises three elements (Sweller, Ayers, & Kalyuga, 2011). The intrinsic cognitive load is the effort associated with the use of specific screens for a specific domain. The extraneous cognitive load is the effort to interact with the screens and associated domain information. The germane cognitive load is the effort to assimilate screen use, domain knowledge and information into long-term memory.

The design of natural user interfaces (NUI) extends the range the interface interactions to include finger wipes, flicks, hand gestures, and body movement. The principle of GUI design facilitates remembering actions, promoting what actions are possible, and how to implement them. Visibility is a fundamental principle of GUI design, and menus make all possible actions discoverable. However, finger wipes, flicks, hand gestures, and body movement are not necessarily uniform between users, cultures, or nationalities (Buxton, 2010; Norman, 2010).

To learn and apply these NUI interactions to a menu driven touch screen was a tacit learning experience that was avoided when developing this system. To do otherwise would be to introduce additional cognitive load (Buxton, 2010; Norman, 2010) on a required but not frequently used system where data quality is the prime concern. As a result, consideration was given to only the basic NUI requirements for this responsive GUI menu-driven responsive design system.

The design brief from the university's senior management contained the human-systems interaction requirement that the software to complete the required tasks should require no formal training. To use software without training implies that a person—any user—must have acquired some use-related knowledge prior to his/her initial use of that software: That is, the user must be able to intuit how to use the software (Sweller et. al., 2011). Research into intuitive design and a user's cognitive load provide the keys to understanding and meeting this requirement.

Another key consideration of use design often given token development time for nonenterprise systems, such as commercial Web sites and smart phone apps, is support of the active user (Carroll & Rosson, 1987). This class of system rarely has help resources embedded; rather, users often post their learning experiences on the Internet. Useful and contextually aligned help resources can aid users in acquiring the information necessary for informed learning and skill development because humans learn from other humans (Rasmussen, 1994; Sweller et al., 2011; Vygotsky, 1978). Humans frequently demonstrate their ability to boot strap, meaning employ self-determined incremental learning, to advance themselves to a required level of knowledge. This ability to apply the acquired depth and breadth of knowledge is a measure of a person's immersion in the domain praxis of a pertinent community of practice (CoP).

Help resources to assist with learning should be designed in line with the best practices of adult learning (Bandura, 1978, 1986, 1997, 2001; Knowles, 1984; Rogers, 2002) and instructional design principles of cognitive load theory (Sweller et al., 2011). When an individual experiences gaps in his/her experiential knowledge and skills that make intuitive use impossible, his/her use of help resources should scaffold his/her learning so that the acquisition of the required knowledge and skills is a positive self-development experience. For example, an experience where video and/or text format help resources instruct and guide the user through a series of learning sequences facilitates the accretion of contextual knowledge and skill. In this way, the help resources impart knowledge and enhance skill by guiding interactive learning experiences that also build self-confidence.

Blackler and Popovic (2015) reviewed the findings of researchers from four continents over the last decade and a half and summarized intuitive interaction research findings and methodologies. These concepts and other current research into intuition and intuitive use were applied in this case study to the selection process for the screen artifacts, their layout, and the visual presentation of information on the screen.

The screens were designed to meet the research specifications for intuitive use and to guide the academic staff through the step-by-step process whereby they edit and monitor a survey throughout its life cycle. The series of screens were designed specifically to scaffold users through the CLSE process without training on how to undertake a CLSE and the university's process for this survey without having seen the CLSE software previously. The objective was to present familiar objects in familiar contexts in a controlled sequence that facilitated the CLSE process with minimum cognitive load. These precepts embody current research into intuition, intuitive design, and cognitive load theory.

The principles applied to the design of the staff screens also were applied to the screens delivering the survey to the students, the survey design, and the development screens used by the system administrator. However, discussion on these latter screens is beyond the scope for this paper.

In the following sections, I present the theoretical rationale used to establish the intuitive design criteria and resultant intuitive use analysis methodology. The rationale I propose reviews HCI design criteria, maps them across intuition criteria, considers how they influence the cognitive load, and then develops of the question set to meet the previously discussed criteria to determine successful intuitive design and intuitive use.

INTUITIVE INTERACTION WITH USER INTERFACES

Intuition is defined as reusing pre-existing experiential and formal knowledge in a novel situation. Preliminary research evidence indicates that, in a novel situation, intuition is a two-stage process based on past experience and experiential knowledge (Blackler & Hurtienne, 2007; Blackler, Popovich, & Mahar, 2009; O'Brien, Rogers, & Fisk, 2010). The first stage involves recognizing a conceptually similar contextual situation previously experienced (Nardi, 1996;

Preece, Rogers, & Sharp, 2007). Blackler et al.'s (2009, p. 1) definition of intuition summarizes the consensus on the essential elements of intuition: "Intuition is a type of cognitive processing that utilizes knowledge gained through prior experience. It is a process that is often fast and is nonconscious, or at least not recallable or verbalizable."

The second stage allows the individual to use the previously acquired skills and knowledge to undertake and successfully complete purposeful actions in the novel situation (Bødker, 1991; Checkland, 1999; Suchman, 1987). Intuitive use of software is the leverage of prior knowledge to reproduce familiar interactions with feasibly familiar objects in a novel context to achieve the anticipated outcomes. The intuitive process attempts to integrate already acquired skills and knowledge with the unfamiliar context by executing the interactions implied by the affordances of the screen objects (Gibson, 1977; Kaptelinin, 2014; Norman, 1998). When the intended intuitive interaction sequence produces the anticipated outcome, the intuitive assessment of the use possibilities is confirmed. Thus, intuition leverages prior knowledge to produce interactions with feasibly familiar objects in a novel context that achieves the anticipated predetermined outcomes.

When the intuitive interaction sequence produces the anticipated outcome, the intuitive assessment of the use possibilities is confirmed. This condition satisfies the requirements of Norman's (1998) "Gulfs of Execution and Evaluation" for successful operation of screen artifacts. Activity theory has a three-level abstraction hierarchy that contextually describes this series of events (Bødker, 1991; Nardi, 1996), and each level of the hierarchy has an associated cognitive load (Blackler, 2006; Embrey, 2015; Rasmussen, 1994; Suchman, 1987). The three levels and associated cognitive loads are

- Activity: A long-term formulation requiring controlled conscious processing of plans for actions or chains of actions that may impose a heavy cognitive load,
- Action: Knowledge application involving controlled conscious processing or rule application requiring intuitive processing with possibly some conscious level of involvement that may impose a moderate or light cognitive load,
- Operation: A scripted and skilled automatic behavior (operationalized action) at a nonconscious level of involvement that may impose a light or negligible cognitive load.

Intuitive Design

Intuitive design is the craft of identifying and implementing familiar screen artifacts and screen layouts in new screen design. Intuitive interaction results from the recognition and successful use of familiar screen artifacts and screen layouts that look and work the way suggested by the familiar artifact affordances. When this goal is not achievable, the design compensates for lack of intuition with well-crafted contextually relevant texts to support self-directed learning experiences.

In their research into models for intuitive design, O'Brien et al. (2010) discussed Blackler's (2006) three principles for intuitive design.

1. Use familiar features from the same domain. Use familiar symbols and/or words, put them in familiar or expected positions, and make the functions comparable with functions users have seen before in similar products that perform the same function. Principle 1 is the simplest level of applying intuitive use.
2. Transfer familiar things from other domains. Make it obvious what less well-known functions will do by using familiar things as metaphors to demonstrate their function.

Principle 2 requires the use of metaphor to transform something completely new into something familiar by relating it to already existing knowledge.

3. Redundancy and constancy: Increase consistency so that the function, location, and appearance of features are consistent among various parts of the design and throughout each part. Principle 3 facilitates users applying the same knowledge and metaphors across all parts of the interface.

From a human–computer interaction (HCI) perspective, Blackler's (2006) principles of intuitive design echo the best practices for interaction design. Elements of the design principles of Shneiderman and Plaisant (2010), Nielsen (1993), and Norman (2004) align with Blackler's (2006) principles, as shown in Table 1. Consequently, it is possible for the user to apply intuitive design without being conscious of it.

When screen design employs familiar screen layouts, artifacts, and terminology, the user's cognitive load is reduced because the actions are readily implemented as operations. Operations, by definition, are skilled behaviors undertaken at a nonconscious level of cognitive processing, resulting in minimal cognitive load. However, due to poor design or operator error, an operation can fail. When an operation fails, the user can reverse engineer the operation.

An operation can be reverse engineered as an action by conceptualizing the skilled behavior. However, this is accompanied by a high extraneous cognitive load, as each step is a consciously controlled decision in a process that requires additional resources from the available but limited resources of working memory. On the completion of this process, germane resources in working memory then process the new information and store it in long term memory. Figure 1 outlines the structure and relationships between working memory resources and cognitive load.

Table 1. The Alignment of Blackler's (2006) Design Principles with Best Practice Design Guidelines by Shneiderman and Plaisant (2010), Nielsen (1993), and Norman (2004).

Principle	**Best Practice Guidelines**		
Blackler	**Shneiderman**	**Nielsen**	**Norman**
Familiarity in the domain	Strive for consistency. Reduce short-term memory load.	Simple and natural dialogue. Speak the user's language.	Use knowledge in the world and knowledge in the head. Map the familiarities accurately. Make things visible, bridge the Gulfs of Execution & Evaluation.
Metaphors from other domains	[not addressed]	Speak the user's language.	Get the metaphor mappings right. Use knowledge in the world and knowledge in the head. Make things visible, bridge the Gulfs of Execution & Evaluation.
Redundancy and consistency	Strive for consistency. Reduce short-term memory load.	Consistency	Use both knowledge in the world and knowledge in the head. When all else fails, standardize.

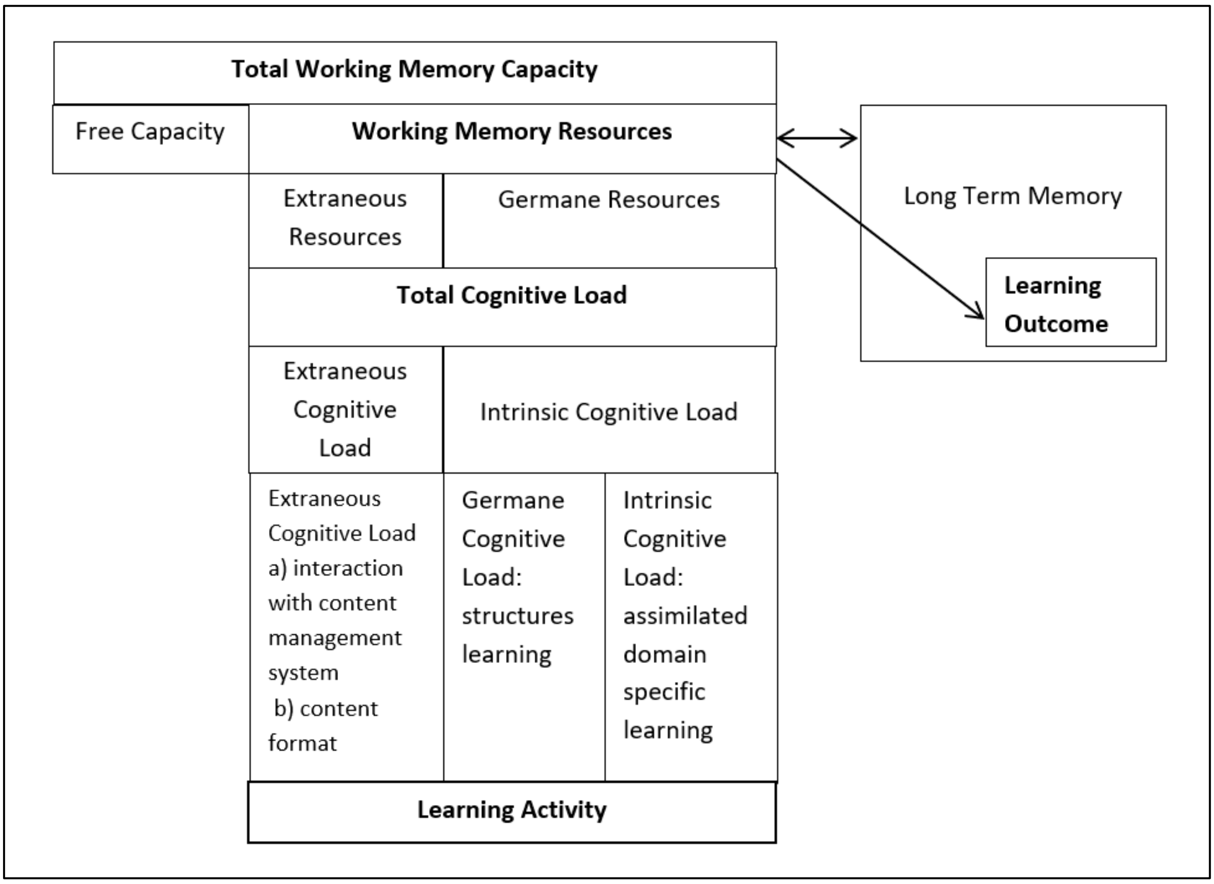


Figure 1. Working memory structure illustrating the relationship between working memory resources and the allocation of those resources to manage the cognitive load associated with a user interface learning experience. The percentage of working memory resources and cognitive load allocated to a particular task vary with each task.

Conversely, to operationalize an action as a skilled behavior requires both extraneous and germane resources in working memory. Once the action is operationalized, the extraneous cognitive load is minimal because the user no longer is required to interpret the artifact or the situated action. The intrinsic cognitive load is also light as the actions and operations are integral to the adopted script for the interaction sequence.

A familiar example of operationalization is learning to drive a manual transmission car. Initially, a driver closely monitors and manages every interaction with the vehicle's controls. The germane resources in working memory are utilized for the overall activity of monitoring the car, controlling one's manipulation of the various systems, and keeping safely on the road. The extraneous resources are utilized to pay attention to all the physical interactions necessary to control the vehicle, specifically, to steer, brake, accelerate, press the clutch, and change gears. As experience is gained and the confidence level in driving skills rises, the vehicle's control interactions become semiautomatous skilled behavior. The extraneous cognitive load correspondingly declines, freeing working memory that can be utilized, if necessary, as germane resources servicing the intrinsic cognitive load of driving the car safely.

Similarly, when an individual is learning how to interact competently with unfamiliar software, the need for additional mental resources in working memory increases the extraneous cognitive load at the expense of the intrinsic cognitive load. This underscores the reason for keeping usability at the forefront in design. If the software cannot be used to obtain the anticipated outcomes, it is impossible to assess whether the software is in fact useful. Usability is the key to determining usefulness, and fit for purpose assessment needs to consider both usability and usefulness.

Usability is a function of the affordances of the screen artifacts. Good usability has its foundation in the system designer knowing the practice of the targeted user cohort (Buxton, 2010). Distributed cognition is the HCI paradigm that best defines what is required for good usability. The affordances of the artifact, which is the vehicle of distributed cognition, are the use-community's praxis crystallized in the artifacts: They transport the history of use and use possibilities across time, location, and the population of the CoP (Lehane, 20012a). These characteristics are the requirements for intuitive interaction.

However, there will be instances of an individual having little or no prior experiential knowledge. In these cases, the help resources of a system should aid in the operationalization of the unfamiliar actions. The help resources should contain explicit step-by-step instructions, obtained from the conceptualization of the operations, to scaffold the user's learning. In this way, the help resources provide a do-as-you-go learning environment that provides partially completed operations as a sequence of procedures to assist the learning and operationalization of the action (Bandura, 1978, 1986, 1997, 2001; Knowles, 1984; Rogers 2002; Sweller et al., 2011). Means to meet this requirement can be presented via, for example, narrated video presentations using screen designs of the system in use and documents that are graphic novelettes with text and images to illustrate the interactions required.

Help resources designed to this paradigm reduce the stress and the cognitive load of learning to use the system. They require fewer extraneous resources in working memory and facilitate more germane resources for the intrinsic cognitive load, if it is required.

Mental Processing

The prerequisite for intuitive use is leverage of prior knowledge with an accompanying skill base and light cognitive load (Buxton, 2010). The context necessary for recognition to trigger possible intuitive interaction is the reproduction of the user's internal knowledge on the screen: the visibility tenet of HCI design. The act of seeing the screen design impacts on the user's mental processes at three levels: visceral, behavioral, and reflective (Norman, 2004).

The lowest of these three levels is the visceral, which responds to sensory stimuli and initiates motor behavior. The visceral is the instinctive processing that instantaneously makes a value judgment on what is good or bad (physically, mentally, and emotionally) and sends appropriate signals to the muscles and to parts of the brain, typically without initial consciousness. The second level is behavioral, which also responds to stimuli and can initiate behavior. The behavioral level can be influenced by the reflective level, whereas the visceral cannot. The highest level is the reflective level, which is isolated from the sensorimotor actuators. It monitors the behavioral and visceral levels and influences behavior.

Affective processing, which includes emotional response, starts with the visceral. Everyday behavior, best described as skilled or operationalized activity without conscious consideration of the activity, is the behavioral level. Finally, the reflective level involves conscious cognition, the contemplative processing of past, present, and proposed experiences.

The cognitive load associated with the use of prior knowledge is a function of the activity's plans, the action levels of mental processing, and the implementation of skilled behavior for the predetermined operations (Buxton, 2010). Rasmussen (1994) classified the different types of information processing required to use "ecological information systems," his term for software programs used in the workplace. The three processing classifications are skill, rule, and knowledge (SRK). Together, they indicate the amount of conscious control a person has over his/her interactions with the system. Blackler's (2006) rationale for the intuitive design criteria and Lehane's (2012a) system acceptance indicator (SAI) were both established using the SRK processing classifications.

Skilled processing is in response to observed signals that trigger highly practiced physical operations that are associated with automatic nonconscious cognitive load. Signals are visceral in that they initiate a low-level automatic response. Signs are visual system characteristics that trigger the mental application of a rule associated with that characteristic or the use of knowledge to generate an appropriate response. The implementation of a rule is intuitive and contextually sensitive. Depending on the context, the rule may be applied nonconsciously or its application may involve some conscious involvement and a light cognitive load. The use of knowledge in response to a sign is a controlled conscious process that can impose a moderate-to-heavy cognitive load, depending on the circumstance. Knowledge work requires reflection to formulate plans based on schema-derived scripts.

Screen design should facilitate the SRK paradigm. Norman's (2004) levels of mental processing and Rasmussen's (1994) information processing hierarchy align mental processing of information with levels of consciousness and behavior. Consideration also needs to be paid to reflective mental processing undertaken as a postbehavior assessment of an activity. This analysis of screen design requirements indicates that reflective assessment needs to consider three knowledge sources:

- knowledge from the skilled and automatic use of signals as transparent symbols;
- knowledge from the intuitive and possibly conscious use of signs as familiar symbols;
- knowledge from the controlled and conscious use of unfamiliar or seemingly familiar symbols.

These knowledge constructs interleave with the operation, action, and activity hierarchical interaction structure of activity theory (Bødker, 1991; Nardi, 1996). Table 2 presents these elements and how they interlink as the requirements for intuitive design of an activity. The extended continuum provides an overview of the relationships between the constructs and their interdependence. To undertake an activity, the objective must be determined from the context of the situation (Buxton, 2010; Checkland, 1999; Suchman, 1987). Consequently, the objective is emergent from the individual's inherent plans and scripts that shape the environment and the activity.

Intuition Elements

The proper selection of screen artifacts and appropriate screen layout from the users' CoP (domain) is essential for intuitive interaction. The computer domain comprises myriad CoPs. An oft quoted but not attributed mantra is "know your user." An extension of this is "know your users'

Table 2. Map of Mental Processing Across the Levels of Cognition, Knowledge, and Context.

Mental processing	**Cognitive processing**	**Level of consciousness**	**Extraneous cognitive load**	**SRK model of task performance**	**Activity theory**
Visceral	Automatic, unconscious engagement	Process is non-conscious	Negligible	Skill–practiced operations with automatic nonconscious cognitive load	Evaluate contextual environment
Behavioral	Intuitive, with some conscious engagement	Possibly some conscious involvement	Light to medium	Rule–nonconsciously applied or some conscious involvement and a light cognitive load.	Contextually triggered scripts for observable behavior
Reflective	Controlled, with fully conscious engagement	Process is conscious	Light	Knowledge–from skilled use and comfortable with use of transparent symbols	Operations - contextually triggered expertise
		Process is conscious	Medium	Knowledge–from intuitive use and comfortable with use of familiar symbols	Actions - contextually scaffolded stages
		Process is conscious	Heavy	Knowledge - from conscious use and comfortable with use of unfamiliar symbols	Activity - conceptually familiar

community of practice" (Lehane, 2012c). A CoP has common practices (praxis) for use and design of its tools. The manner in which the community experiences its practices is explained by the HCI paradigm of distributed cognition (Hutchins, 1995, 2000) and is crystallized in the affordances of the artifacts used. The affordances propose the artifact use possibilities and the methods of use. Identifying the artifacts and understanding the use practice provides insight into the why that is fundamental to the what, when, where, and how of screen and system design (Lehane, 2012a, 2012b, 2012c).

The intuitive interaction continuum assimilates Blackler's (2006) three principles for intuitive interaction:

- use familiar features from the same domain;
- transfer familiar things from other domains;
- [implement] redundancy and internal consistency.

Designing with the implicit HCI intuitive guidelines and Blackler's principles as the primary focus brings as many as possible already known elements into the design. The other HCI guidelines and the functionality to support the required tasks are a secondary focus implemented as appropriate to the stage in the design progress.

The primary focus creates a context that generates a positive visceral response. The secondary focus provides rules and onscreen knowledge for successful interactions based on behavioral knowledge and existing expertise and experience.

Assess Intuitive Design

The research into intuition identified three essential elements as being necessary for the emergence of intuitive interactions.

1. A context that presents visual triggers for prior knowledge and expertise to be expeditiously and nonconsciously accessed prior to initiating a scripted behavior for a particular action and subsequent operations.
2. Initiating the interactions scripted by the suggested familiarity of the novel situation.
3. Completion of the interactions delivering the outcomes suggested by the familiarity, thereby confirming the successful leverage of prior knowledge.

The requirements for intuitive use have been identified and discussed. However, the question arises as to how intuitive design can be evaluated, along with the corollary of how to use the feedback to improve both the design and the design process. Anecdotal evidence alone makes it difficult to validate and verify. It would be advantageous to have an instrument that could be readily applied with consistent reproducible results.

The unified theory of acceptance and use of technology (UTAUT; Venkatesh, Morris, Davis, & Davis, 2003) integrated technology acceptance and use concepts. Previous work on the evaluation of design produced the SAI (Lehane, 2012a). The development of the SAI mapped the SAI questions to the UTAUT question sets and theoretic frameworks (Lehane 2008; Lehane & Huf, 2007). The SAI is a validated survey instrument that can be used to evaluate system use within an HCI designed technology acceptance framework. Previously, Blackler's (2006) three principles of intuitive interaction were shown to align with existing principles for HCI design. Therefore, Blackler's principles should be able to be evaluated using the SAI.

The SAI is a survey instrument of 25 statements that are assessed using a 5-point Likert agreement scale. The responses are mapped to usability engineering concepts of system acceptance, usefulness, functionality, utility, and usability. This is achieved by mapping the elemental properties of active user, distributed cognition, affordances, immersion in the CoP and the context of the work environment onto the usability engineering concepts (Lehane, 2010). The SAI question set was reviewed and six questions were considered to form a concise subset that describes the developed intuitive mental processing schema. Table 3 presents the SAI questions, their UTAUT determinant, and the theoretical basis for the inclusion of the question in the question set.

The mental processing and screen design aspects associated with each question are presented in Table 4. This table outlines the suitability of each question for evaluating the associated mental processing.

A comprehensive reflective assessment of an activity considers all three mental processing activities and the predominant screen design aspect impacting on the mental processing. Thus, the GUI designer considers the initial visual impact. The DMI/NUI aspects are considered for the perception of the on-screen affordances, interaction inputs for an outcome, and the outcome itself. Visceral and behavioral processing are separate stages but, by its nature, reflective assessment must contemplate the skills and behaviors as well the overall activity. The question set should facilitate this and be able to assess the intuitive qualities of the screen design, screen layout and the interaction design:

Table 3. SAI Questions as a Statement for Agreement with its Associated UTAUT Determinate and Technology Acceptance Theory.

SAI Questions	UTAUT Determinate	Technology Acceptance Theory
Q1. The screen for this software looks like other screens I have used.	Effort expectancy: perceived ease of use (Key concept: skillful)	Technology Acceptance Model[1] & Innovation Diffusion Theory[2]
Q2. When I look at the icons and menus, etc., on the screen I know what to use and how to use it.	Effort expectancy: perceived behavioral control/ease of use (Key concept: clear & understandable)	Technology Acceptance Model & Innovation Diffusion Theory
Q3. For the things that I use, this software looks and works the same way every time.	Effort expectancy: perceived behavioral control/ease of use (Key concept: software does what I want it to do)	Technology Acceptance Model & Innovation Diffusion Theory
Q4. There is always enough information on the screen when it is needed.	Facilitating conditions – perceived behavioral control/perceived usefulness (Key concept: flexible)	Technology Acceptance Model & Theory of Planned Behavior[3]
Q5. I think I will be able to use this software without asking for help from the experts who know how to use it.	Facilitating conditions: perceived behavioral control/ self-efficacy (Key concept: knowledge)	Technology Acceptance Model & Theory of Planned Behavior & Social Cognitive Theory[4]
Q6. I would recommend this software to a colleague or friend.	Social influence: social factors (Key concept: recommend)	Model of PC utilization[5] & Social Cognitive Theory

Notes. 1. Models how uses accept and use technology (Davis, as cited in Venkatesch et. al., 2003). 2. Rational for the rate of dispersion of new ideas and technology (Rogers, as cited in Venkatesch et. al., 2003). 3. Explain behaviors over which people can exert self-control (Mathieson, as cited in Venkatesch et. al., 2003). 4. Knowledge acquisition occurs in the context of social interactions (Bandura, as cited in Venkatesch et. al., 2003). 5. Knowledge acquisition occurs in the context of social and technological interactions (Compeau & Higgins, as cited in Venkatesch et. al., 2003).

Table 4. Mental Processing and Associated Questions Relating to Screen Design Aspects.

Mental Processing	SAI Questions	Screen Design
Visceral—influenced by skill	The screen for this software looks like other screens I have used.	GUI aspects
Behavioral—influenced by rules	When I look at the icons and menus, etc., on the screen I know what to use and how to use them.	DMI/NUI aspects
Reflective—influenced by skill knowledge	For the things that I use, this software looks and works the same way, every time.	GUI and DMI/NUI aspects
Reflective—influenced by action knowledge	There is always enough information on the screen when it's needed.	GUI and DMI/NUI aspects
Reflective—influenced by meta knowledge	I think I will be able to use this software without asking for help from the experts who know how to use it.	GUI and DMI/NUI aspects
Emotional	I would recommend this software to a colleague or friend.	Response that assigns value to objects or events

Note. GUI is the graphical user interface; DMI/NUI is direct manipulation interface/natural user interface.

- Question 1 initial reaction to the user interface;
- Question 2 confirms Question 1;
- Question 3 confirms Question 2;
- Question 4 confirms findability and usefulness for an active CoP member;
- Question 5 confirms the design uses the practice of the user's CoP;
- Question 6 confirms the previous question set with an emotional personal association by the survey participant (ownership).

The previous sections discussed the characteristics of intuition and identified three elements required for intuitive interaction. The three elements are familiar user expectations, confident interactions, and the successful leverage of prior learning. The SAI intuitive use subset of questions is pertinent to these concepts.

An assessment instrument must allow results to be interpreted consistently over a diversity of projects. Table 5 provides a summation table that associates each question of the SAI on screen design with an activity theory concept, the SRK model for information processing, and Norman's mental process. A low response for an individual question indicates that that element of the design criteria requires attention by way of fieldwork and/or design review.

A high level of agreement demonstrates that the screen design followed the practice of the users' CoP and thus the extraneous cognitive load of users will be low. In such a case, the system allowed the users to draw on their extant knowledge and skills for intuitive interaction with the visual, graphic, interaction, navigation, and information elements of user interface.

The question set was used to evaluate the intuitive use of a new software system at a regional university in Australia. I present the case study in the next section.

Table 2. System Design Criteria Applicable to the SAI Statements for Agreement.

SAI Question	Design Criteria	Activity Theory	SRK MODEL	Mental Process
The screen for this software looks like other screens I have used.	Usability, distributed cognition, affordances	Evaluate contextual environment	Skill—sign	Visceral
When I look at the icons and menus, etc., on the screen I know what to use and how to use them.	Usability, distributed cognition, affordances	Contextually triggered scripts	Rule—signals	Behavioral
For the things that I use, this software looks and works the same way, every time.	Usability, distributed cognition, affordances	Operations - contextually triggered expertise	Knowledge from skilled use; comfortable with use of transparent symbols	Reflective
There is always enough information on the screen when it's needed.	Usability, distributed cognition, affordances	Actions contextually scaffold stages	Knowledge from intuitive use; comfortable with use of familiar symbols	Reflective
I think I will be able to use this software without asking for help from the experts who know how to use it.	Depth of immersion in CoP	Activity within conceptually familiar schemas & scripts	Knowledge from conscious use; comfortable with use of unfamiliar symbols	Reflective

THE CASE STUDY

In this section, I discuss the design and development of the academic staff interfaces presented in the Introduction. The discussion focuses on the screen and interaction design to meet the staff user requirement that the system enables users to complete the required tasks without formal training. The overall design of the system and user interfaces for all user classes is out of scope for this paper. However, a paper discussing the design for the screens, interactions, and user experience is proposed.

The preamble on intuitive design and interaction outlines the theoretical considerations given to the design of all the system's user interfaces (UIs). The HCI design paradigm used for the development process was the spiral life cycle model (Boehm as cited in Preece et al., 2007; Winston, 1970) applying user-centric rapid application development (Millington & Stapleton as cited in Preece et al., 2007).

Progress reviews evaluated the work by using Checkland's (1999) five measures of performance:

- efficacy—successful in producing the required output
- efficiency—the minimum resources are used to obtain the output
- effectiveness—the processes modeled are indispensable
- ethics—the process has moral integrity
- elegance—the overall design and its elements are aesthetically pleasing.

There is a caveat to elegance in that the aesthetics of graphic design must never take precedence over usability.

The programming for the CLSE system was based on a survey life cycle (SLC) that incorporated, but was not limited to, functionality for the design and development of the survey question set, managing the opening and closing dates for the stages of the SLC, changing the data sets displayed with each life cycle stage, controlling security and access to personal and corporate data, ensuring that legacy question sets could not be changed after a survey was closed, and generating reports. The SLC programming also included control of progress through the SLC and the CLSE processes.

The brief required that the academic staff interfaces were to be accessible on devices ranging from notebooks to desktop computers and use the dominate browsers for Microsoft, Android, and Apple operating systems. To meet this platform specification, a responsive design was the goal, rather than developing a separate application for the mobile solutions. Mobile phone platforms with their small screens and onscreen keyboards were excluded from the brief due to their potential to cause usability issues in reading and inputting data.

The design brief was quintessentially short: A CLSE system that the academic staff could interact with without training in the use of the software or training in their role for course-level student evaluation. Over a period of years, while running training workshops, the university's course evaluation and survey officer had identified and recorded shortfalls in functionality, on-screen information, and software use issues. Consequently, the new system's functional requirements for academic staff were documented in detail. This evidence-based research was invaluable as it established the usability and functional requirements.

The design process took place within three sequential but overlapping phases—nominal, descriptive, and formative—in line with ecological interface design principles (Vicente, 1999). The nominal phase analyzed the current system and reviewed the already documented evidence-based requirements; the descriptive phase investigated the contextual-use requirements of the current system and the current shortfalls. The formative phase encompassed the design and development phases undertaken after detailing the use requirements and the desired outcomes of the future system. The formative phase also incorporated the initial design of the required outcomes, the processes to deliver the new UIs, the associated interaction designs, and the coding of the business rules and screens.

Gestalt and aesthetic considerations proved important in the design and presentation of the information on the screen. With placement, text, and color choices, the graphic designer improved the readability of important but previously mundane notices. Quality graphic art represents an important aspect of screen design because positive initial visual impressions are critical for usability and system acceptance. However, aesthetics must never take precedence over usability: If graphic design dictates that an object should be on the right side of the screen but usability precedence dictates that users expect the object to be on the left, then the object must be placed on the left.

Academic personas were developed to identify the predominate subsets of staff user characteristics, academic roles, IT experience, and attitude toward the various technologies used. The personas grounded the use cases with evidence-based issues that identified the pros and cons of existing and proposed functionality and screen layout. The selection of screen artifacts for interaction controls, such as icons and terminology, was drawn from the Microsoft pallet because the university specifies Microsoft products for use by staff and students. Apple products are supported but not specified for staff and student use.

Academic subject matter experts (SMEs) were brought into the process early in the development of each set of screens that supported an individual activity. The selected SMEs possessed attributes identified in the purposive sampling matrix that was developed from the personas. Various levels of expertise and experience were sought in areas of academic role, familiarity with CLSEs, familiarity with the previous CLSE software, IT skills, and English-as-a-second-language experience.

Changes to the layout and changes to the academic terminology were based on the recommendations of the academic SMEs. Where the SMEs had difficulty in using the software or understanding the terminology, designers modified the software to fit their mental models. The SMEs first exposure to a design was to step through the proposed interaction design on a wireframe. Feedback was sought on all aspects of screen design: the on-screen artifacts, screen layout, terminology, and the sequence of the steps to progress through the process, as presented on the wireframe.

The first screen was produced after the initial mock-up and test. The SMEs were brought back to review subsequent versions of the screens, the interaction design, and the help resources. Multiple reviews resulted in identified issues requiring changes to the screen design and the help resources. The help resources of the case study were designed using the best practice adult learning (Bandura, 1978, 1986, 1997, 2001; Knowles, 1984; Rogers, 2002) and instructional design principles of cognitive load theory (Sweller et al., 2011).

The screen reviews, on occasion, identified places in the process where the interaction design needed to constrain the available branching options to ensure the process proceeded as required. In other cases, the interaction design required modifications when a screen had to be expanded to two screens to accommodate the process or when two screens had to be condensed into one to constrain the process.

Information relating to various aspects of software use or information about the CLSE process and role specific responsibilities were allocated separate areas on the screen. The location, area allocated, and color reflected best practice usability and the relative importance of the information in the CLSE process. Gestalt and aesthetic principles were applied to the overall layout, the design of sections, and the allocation of fonts, line spacing, and color. The location, shape, and color of icons and the design of other on-screen artifacts for monitoring and control were tested by experienced and novice computer users to ensure a base level of use-familiarity and outcome expectations.

The on-screen information and the page-specific help resources were designed and developed to scaffold the academic staff through their assigned activities. The help function used multimedia and text formats. The videos were produced to enhance the learning experience for staff members without experiential knowledge of CLSE or the survey editing process and to assist nonnative users of English.

During design discussions, the various activities associated with the tasks were identified. The various activities were termed “what brings them to the screen.” In this perspective, “them” encompassed all the proposed classes of users of the CLSE system. This allowed for consideration of the inputs and outcomes of each activity, its associated functional requirements and related data fields, as well as the interactions to monitor and control each activity of the task.

By way of example, a number of activities can bring a staff member to the screen. For instance, a staff member can access the software to review a survey about to be run, monitor the response rate of an open survey, or provide examiner feedback. Each stage of the survey life cycle brings implications for the staff, the students, and the system administrator. The set of screens designed for each audience at each stage underwent the same rigor of design outlined in this section. An overview of the staff screens developed for the CSLE system is presented in the following section of the paper.

STAFF USER INTERFACES

In this section, the discussion of staff UIs covers intuitive use in more detail. However, stepping through the interaction design for each screen to assess the user experience is the focus of a separate user experience paper under consideration.

The staff UIs were developed from the scenarios of situated actions (Buxton, 2010; Carroll, 2003; Suchman, 1987). The screens and interactions were designed to support the academic staff in the following activities.

The teaching staff role involves

- reallocating the examiner role if he/she is not the course’s examiner;
- checking that all courses are listed, and if not, contacting the examiner to be added to the staff for the course;
- making sure all staff for evaluation are listed by name, if this feature is offered for the survey;
- previewing the survey, checking any questions that have been added to the survey by the examiner, and discussing options for additional questions from the approved question list with the examiner; and
- reviewing the evaluation to see what the students are going to see.

An examiner’s editing role entails

- checking that all his/her allocated examiner courses are listed;
- checking that all staff teaching the course are attached to the course;

- removing any staff listed who are not teaching the course;
- allocating staff for evaluation, if this feature is offered for the survey (students will see the staff member's name);
- adding questions from the list of approved questions, if appropriate;
- reviewing the evaluation to clarify what the students are going to see;
- monitoring the status of the survey;
- monitoring response rates; and
- accessing reports and provide examiner feedback as appropriate.

After selecting the current survey campaign on the staff landing page, the courses in that survey iteration for which the staff member is either examiner or teacher are presented in a list. A segment of this course list is presented in Figure 2. Course description, academic role, and SLC status are displayed for monitoring the courses progress throughout the SLC.

Specific courses in a person's list can be found using the search function; a course that has not been associated with a person also can be accessed using the search function. The system's search functionality supports search, filtering, and sorting all data contextually relevant to that screen. Selection of the Edit button opens the course for editing. Even after an initial editing, a course survey can be modified while the survey campaign is open for editing. Teaching staff who are not the examiner can preview the survey by selecting the Preview button.

Teaching staff also can view the pool questions from the preview. Pool questions are additional questions that the survey owner has the option to ask. Access to the pool questions allows the teaching staff and the examiner who may be on different campuses to simultaneously view and

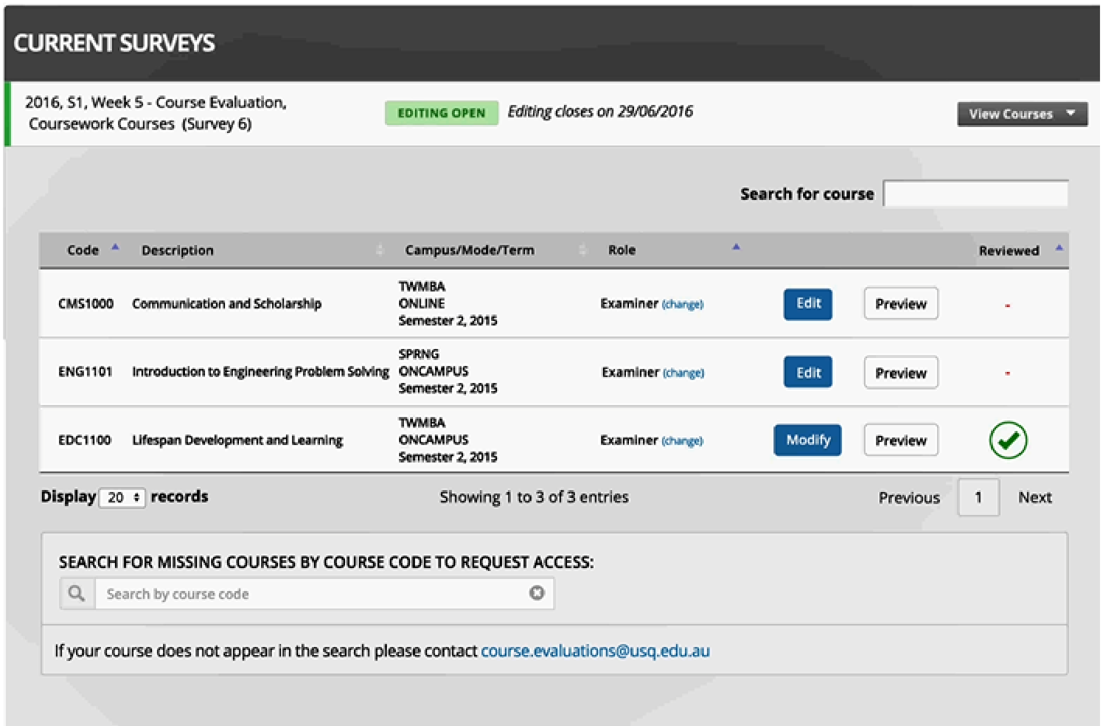


Figure 2. Academic staff view of the staff landing page that lists the surveys currently running for courses in which the academic staff member has either examiner or teaching responsibilities.

discuss possible additional questions conveniently over the phone. This helps to shorten the timeline for survey development. An edited survey has a green tick in the Reviewed column and is available for modification. The Reviewed status is useful for monitoring the completion rate of the editing process. Once a survey is open to the students, the Reviewed column data are replaced with a column showing the student response rate as a percentage and tally.

The Editing Process

The editing process was developed as a five-step procedure with individual screens to scaffold the user through the process:

- confirm examiner and teaching staff;
- set up staff evaluation (if part of the survey);
- review questions and add questions, if appropriate;
- preview the survey;
- mark as reviewed.

Progress through the editing process is monitored by the course review progress bar. The icons for each stage in the process are links to the screen for that stage. As each stage is completed and saved, a green tick replaces the stage number in the icon. Video and text help resources are available, as well as a Back button for navigation.

The first editing screen, as shown in Figure 3, asks the fundamental question of course ownership to establish/confirm examiner rights or teaching staff only access. The development of this functionality to control access to the survey and examiner rights allowed the academic

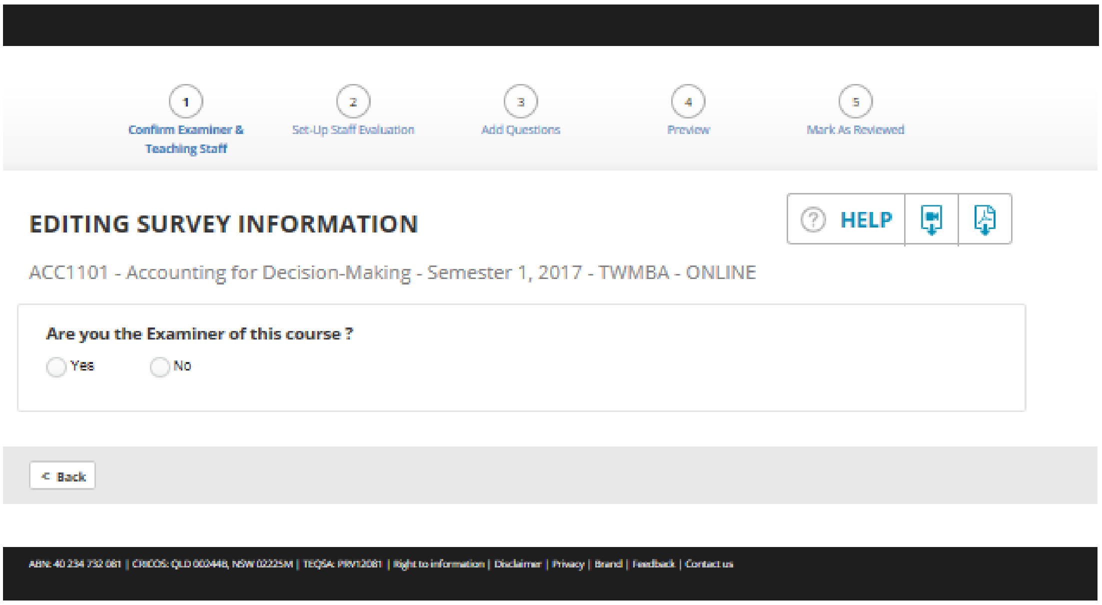


Figure 3. The Examiner Rights screen allows academic staff to establish their relationship to the course and the survey. An examiner has privileges for editing survey that a nonexaminer, teaching only, staff member does not.

staff to make any required changes to staffing without assistance from the system administrator, thereby empowering academic staff.

A "No" response to the examiner rights opens an overlay that, on completion, initiates an email to the new examiner. The previous examiner is then reassigned to a teaching staff role and returned to the landing page. A "Yes" response calls the next screen to confirm the teaching staff.

The screen to confirm teaching staff, Figure 4, has the same visual appearance as the first screen in the process: progress bar, page-specific video, and text help. Additional CLSE information is presented on the screen above the associated data fields and control artifacts.

Academic staff associated with the course in the university's academic database are the default teaching staff for the course. The examiner can remove the default-entered staff if they are no longer associated with the course, and new staff can be added using the search function.

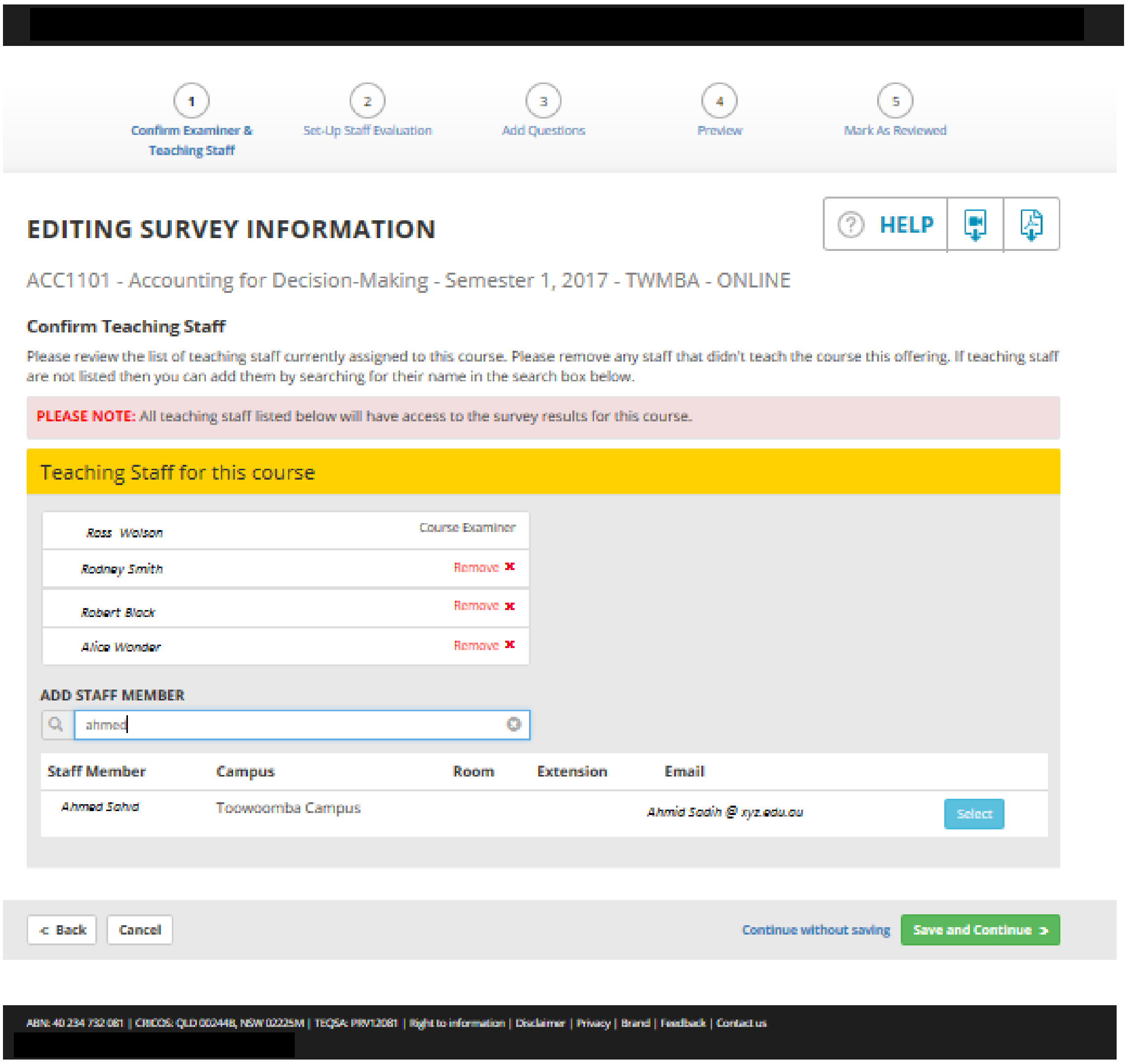


Figure 4. The Confirm Teaching Staff screen is a responsibility of the course examiner after his/her role has been confirmed on the previous screen. This screen facilitates the addition or removal of teaching staff (fictitious names in the figure) from the list of staff who can access the survey and the survey reports.

The interactions to search for and select new staff are guided by the use of familiar screen artifacts, text fields, icons, and on-screen instructions.

Information relating to the process is presented first as an overview to provide context. Then the actions in the required sequence for this stage of the CLSE process are identified, along with the controls for each action. Working with the document takes the reader through the cognitive and behavioral procedures necessary to complete the process. Steps that are not applicable can be skipped. The last instruction explains how to proceed to the next stage of the CLSE process.

The page-sensitive help resources for the system were developed to instruct and guide a person requiring assistance to engage with the necessary interactions and to explain the importance of each stage in the CLSE process. Thus, in this context, the Confirm Teaching Staff help pdf (see Figure 5) is presented to explain the help document format.

The third step in the process is the editing of the survey itself. The visual appearance and navigation are unchanged. However, the second stage icon will indicate that stage has been completed and saved, as demonstrated in Figure 6.

The survey questions are grouped by evaluation type and response rating scale. The questions are either core for the survey or optional questions from the pool. The inclusion of pool questions is at the discretion of the examiner and can be added or removed at any time prior to the survey being released.

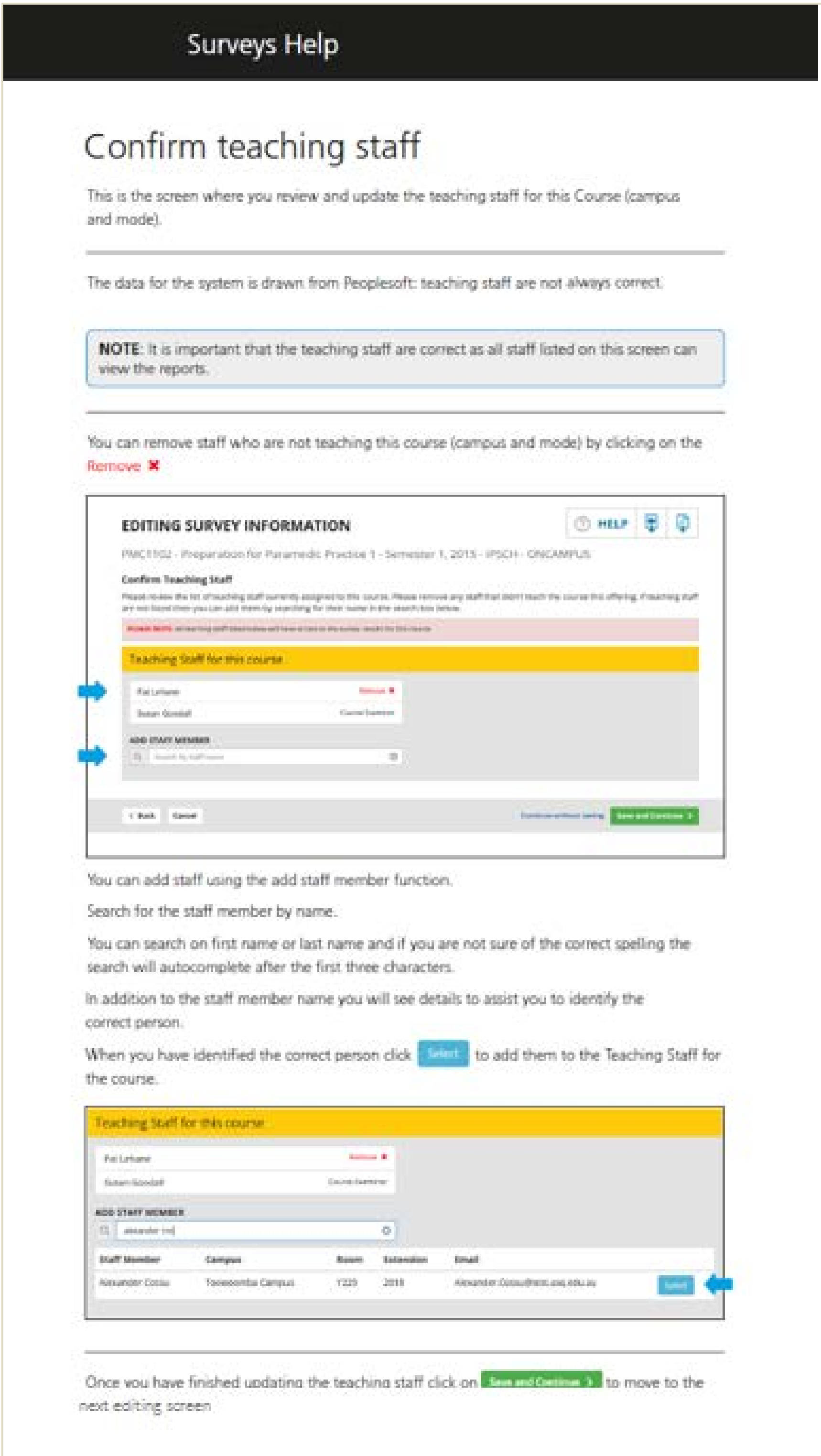
The final stage of the editing process is to review the survey. This screen, shown in Figure 7, allows the examiner and teaching staff to see the survey as the students will see it. The Preview Student View screen also allows the teaching staff access to the pool questions. Only the examiner can add pool questions to the survey but teaching staff can review the pool questions available and make requests to the examiner to have pool questions added.

The report section closes the CLSE process. The reports are open to staff and students after they have been checked by the system administrator to ensure there are no system errors affecting the results. The reports provide each student with a copy of their individual responses and a nonidentifiable aggregated summation of their cohorts' responses and comments. Staff can access nonidentifiable aggregated summations of class responses and comments.

The report, shown in Figure 8, presents the data in a series of formats: mean comparison or response frequency with vertical or horizontal graphs. The viewer can select the desired presentation format. The extensive detail of the quantitative data analysis in not reported in this paper. The help resources explain in detail the rationale for the statistics used and the interpretation of the aggregated data.

Student comments on the courses are also displayed in the reports and provision is made for examiners to provide feedback to current students and the next cohort taking that course. The provision of examiner feedback completes the communications loop of (a) asking the initial question, (b) receiving a response and any additional comments, and (c) providing feedback to the additional comments.

The screens discussed illustrate how the intuitive guidelines were implemented in the context of supporting the academic staff to manage their role and associated responsibilities in a survey campaign.



Surveys Help

Confirm teaching staff

This is the screen where you review and update the teaching staff for this Course (campus and mode).

The data for the system is drawn from Peoplesoft: teaching staff are not always correct.

NOTE: It is important that the teaching staff are correct as all staff listed on this screen can view the reports.

You can remove staff who are not teaching this course (campus and mode) by clicking on the Remove ✖

You can add staff using the add staff member function.

Search for the staff member by name.

You can search on first name or last name and if you are not sure of the correct spelling the search will autocomplete after the first three characters.

In addition to the staff member name you will see details to assist you to identify the correct person.

When you have identified the correct person click Select to add them to the Teaching Staff for the course.

Once you have finished updating the teaching staff click on Save and Continue › to move to the next editing screen

Figure 5. The Confirm Teaching Staff Help pdf is text format help resource that provides a step by guide to adding or removing a staff member from the list of people who can access the survey and the survey reports.

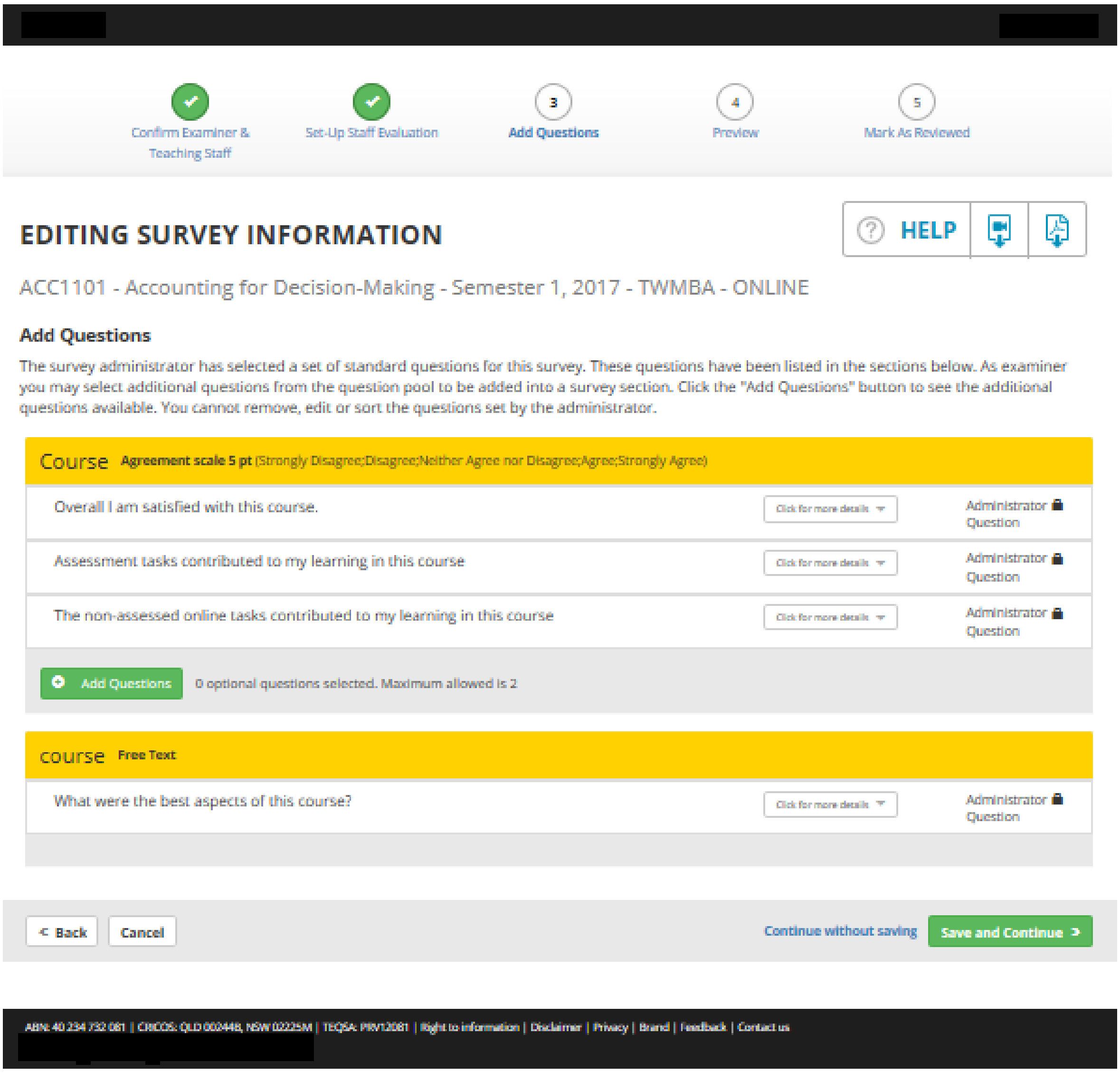


Figure 6. The Review the Survey Questions screen presents the questions proposed for the survey. The questions are grouped by response format. For example, all 5-point agreement scale questions are shown together and separate from the free-text questions. The examiner can add additional questions from a prepared secondary question set.

INTUITIVE USE ANALYSIS

The six items from the SAI question set selected for *intuitive use evaluation* were discussed and presented earlier in the paper. The SAI is a validated survey instrument for assessing system acceptance using models, theories, and frameworks from the discipline of HCI. The intuitive use subset was determined from current research into intuition, where intuition is defined as reusing pre-existing experiential and formal knowledge in a novel situation.

Prior to using the survey, ethics approval was sought from and given by the University's Ethics Committee. An email with a link to the survey was sent to all academic staff who had

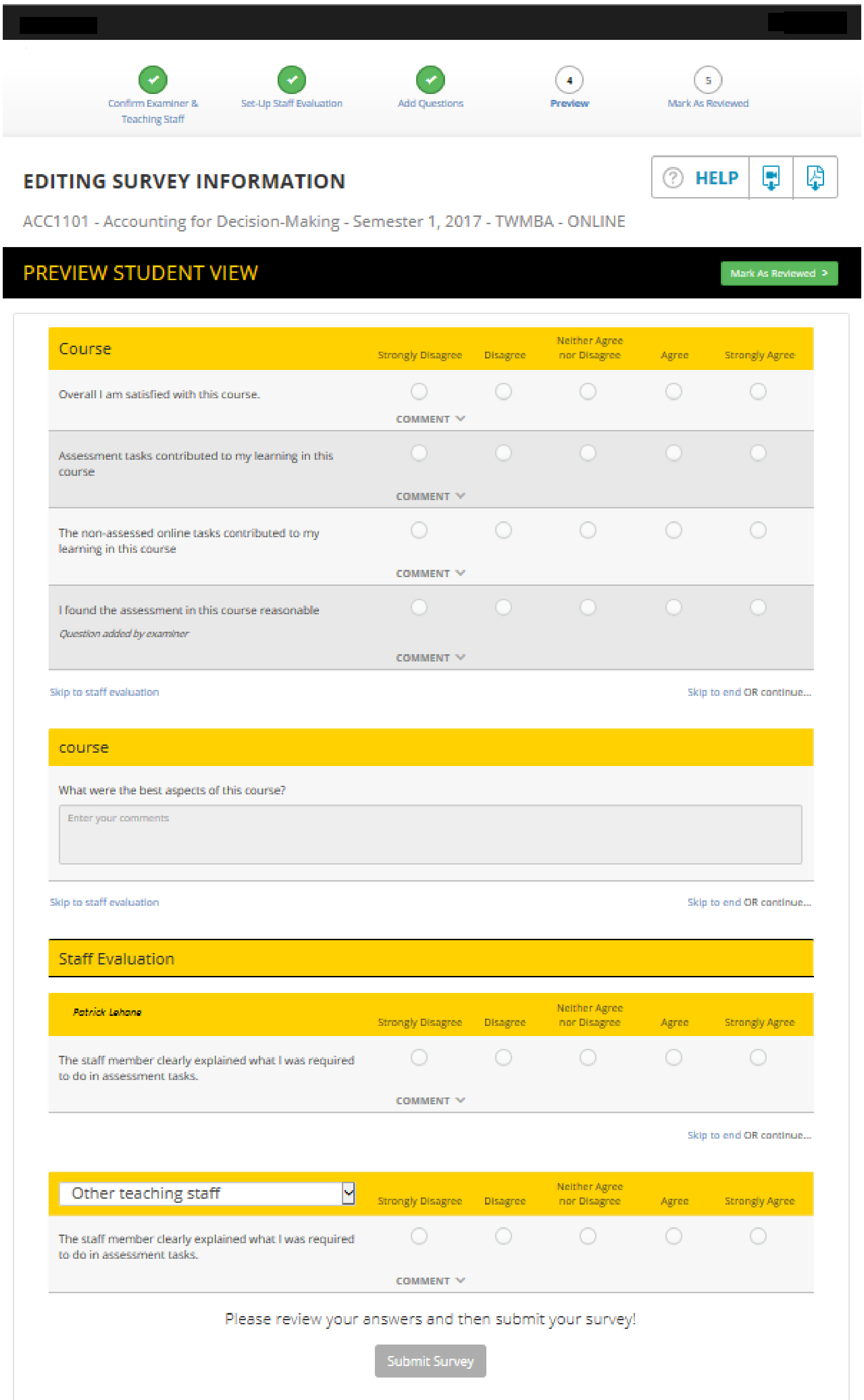
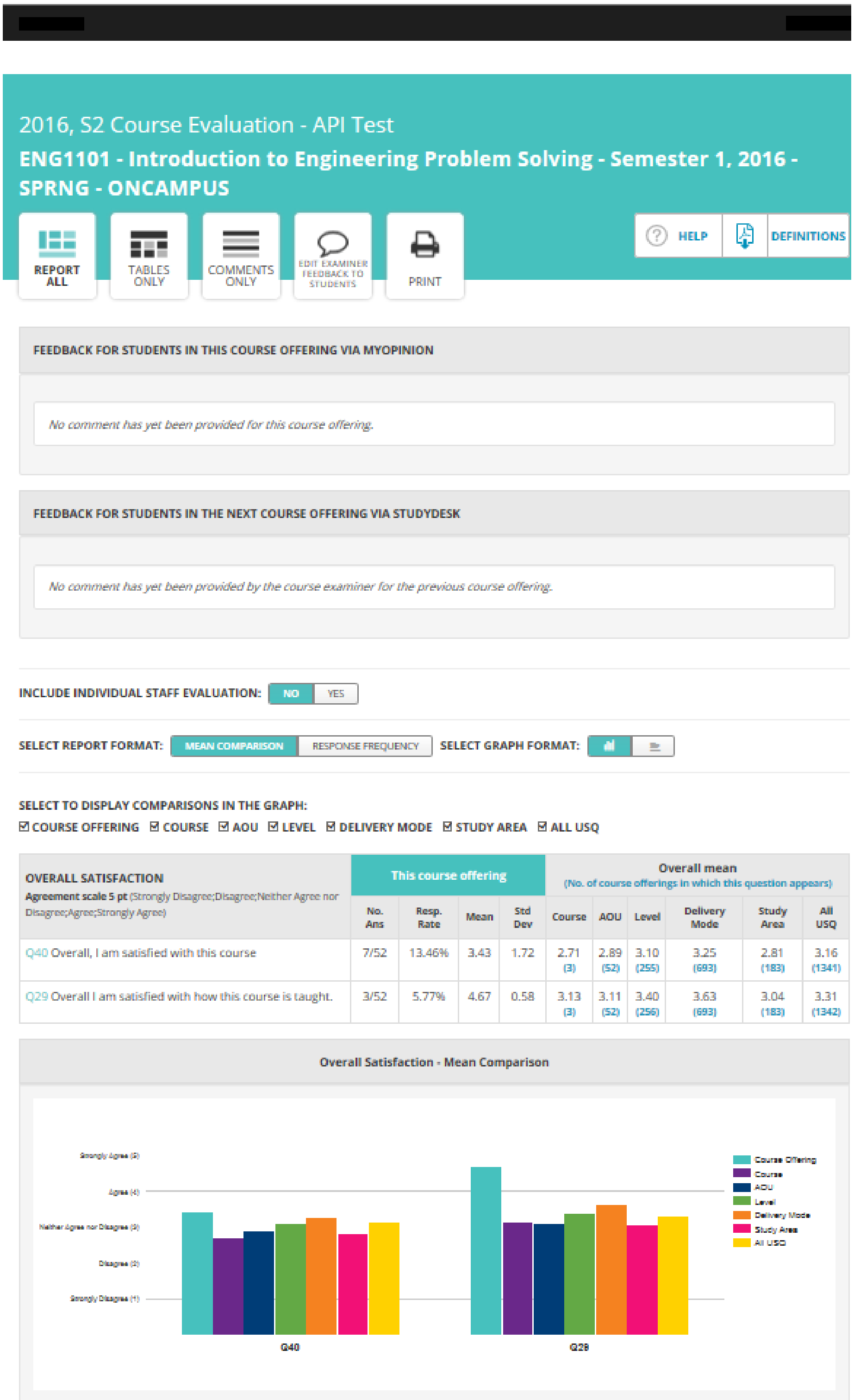


Figure 7. The staff Preview Student Survey screen presents the staff member with the same view of the survey that the student will see. This allows the examiner to consider changes to the survey and observe how those changes will influence the look and feel of the survey.



[redacted] [redacted]

2016, S2 Course Evaluation - API Test

ENG1101 - Introduction to Engineering Problem Solving - Semester 1, 2016 - SPRNG - ONCAMPUS

REPORT ALL | TABLES ONLY | COMMENTS ONLY | EDIT EXAMINER FEEDBACK TO STUDENTS | PRINT | HELP | DEFINITIONS

FEEDBACK FOR STUDENTS IN THIS COURSE OFFERING VIA MYOPINION

No comment has yet been provided for this course offering.

FEEDBACK FOR STUDENTS IN THE NEXT COURSE OFFERING VIA STUDYDESK

No comment has yet been provided by the course examiner for the previous course offering.

INCLUDE INDIVIDUAL STAFF EVALUATION: NO YES

SELECT REPORT FORMAT: MEAN COMPARISON RESPONSE FREQUENCY SELECT GRAPH FORMAT:

SELECT TO DISPLAY COMPARISONS IN THE GRAPH:
☑ COURSE OFFERING ☑ COURSE ☑ AOU ☑ LEVEL ☑ DELIVERY MODE ☑ STUDY AREA ☑ ALL USQ

OVERALL SATISFACTION **Agreement scale 5 pt** (Strongly Disagree;Disagree;Neither Agree nor Disagree;Agree;Strongly Agree)	This course offering				Overall mean (No. of course offerings in which this question appears)					
	No. Ans	Resp. Rate	Mean	Std Dev	Course	AOU	Level	Delivery Mode	Study Area	All USQ
Q40 Overall, I am satisfied with this course	7/52	13.46%	3.43	1.72	2.71 (3)	2.89 (52)	3.10 (255)	3.25 (693)	2.81 (183)	3.16 (1341)
Q29 Overall I am satisfied with how this course is taught.	3/52	5.77%	4.67	0.58	3.13 (3)	3.11 (52)	3.40 (256)	3.63 (693)	3.04 (183)	3.31 (1342)

Overall Satisfaction - Mean Comparison

Figure 8. The survey reports are extensive. This figure displays the top section of a final report. Student comments and academic staff responses to those comments are available, as well as descriptive statistics for the survey participants' responses.

professional responsibilities to access the CLSE system. There were no lead-in or follow-up emails. This convenience sampling method contacted 517 academic staff who had teaching and or examiner responsibilities. There were 399 examiners listed.

The survey received 75 responses, 63 responses were from academic staff who used the CLSE system and 12 from academic staff who indicated that they had not used the CLSE system and gave no explanation as to why not. One participant did not answer any questions other than to claim to have used the CLSE system. Of the remaining 62 participants, 58 responded to all six items relating to their use experience. However, there were only 54 examiner responses. Academic staff with only teaching responsibilities did not use the full functionality available to academic staff and their responses were not included in the data for analysis. Consequently, just the examiner subset of responses was analyzed because their individual data sets covered all interaction aspects of CLSE system use necessary for quantitative analysis, thereby ensuring that correlations between the data items were not biased by unequal weightings or dependence.

The number of examiner responses suitable for quantitative analysis was 54, reducing the response rate to 13.5%. Research by Nulty (2008) indicated that this response rate for a population of this size meets conditions for 10% sampling error and an 80% confidence level, an acceptable rationale to test the data for compliance with conditions for quantitative analysis.

As well as the six questions on the user experience, the survey contained questions relating to the survey participants' prior experiences in course-level evaluation, their individual roles and workloads, the tasks undertaken using the software, and the participants' information technology skill sets. The criteria for assessing experience and information technology skills are presented in Appendix A and are presented to ensure the transparency of the research.

These demographic questions facilitated my identifying factors that may influence large population survey data in an analysis using the full SAI question set (Venkatesh et al., 2003; Lehane, 2008, 2012a). Possible demographic influences on the participant's responses are not discussed in this report because of the small data set.

Human-Computer Interaction Analysis

The SAI calculates the descriptive statistics and then presents the mean value for each question as a plot. In Figure 9, this plot is called Series 1. The intuitive use concepts associated with each question is also presented on the graph; for example, Question 1 relates to familiar screen look and feel. The SAI value range is 0 to 4; a score of 2.5 or greater is indicative of strong user affirmation of an SAI concept or item. A score between 1.5 and 2.5 is affirmation of a concept or item, while a score less than 1.5 indicates that investigative fieldwork should be carried out to identify the issues underlying the low response.

The SAI value of 3.2 for the intuitive use concept is indicative of consistently strong user-affirmations that the design of the user interface and the design of the user interactions facilitate an intuitive user experience. Subsequently, the graphic elements of the GUI, the affordances of the controls of the DMI/NUI, and familiar and confident progress through the required activities may be considered to for analysis be due to familiar screen artifacts that were operated as expected and with the resultant anticipated outcomes.

The CLSE system was designed to facilitate the use of pre-existing experiential and formally acquired computer use knowledge. Where possible gaps existed between a user's knowledge and the needs of the system, online training was provided by the help subsystem. Ultimately, the overall

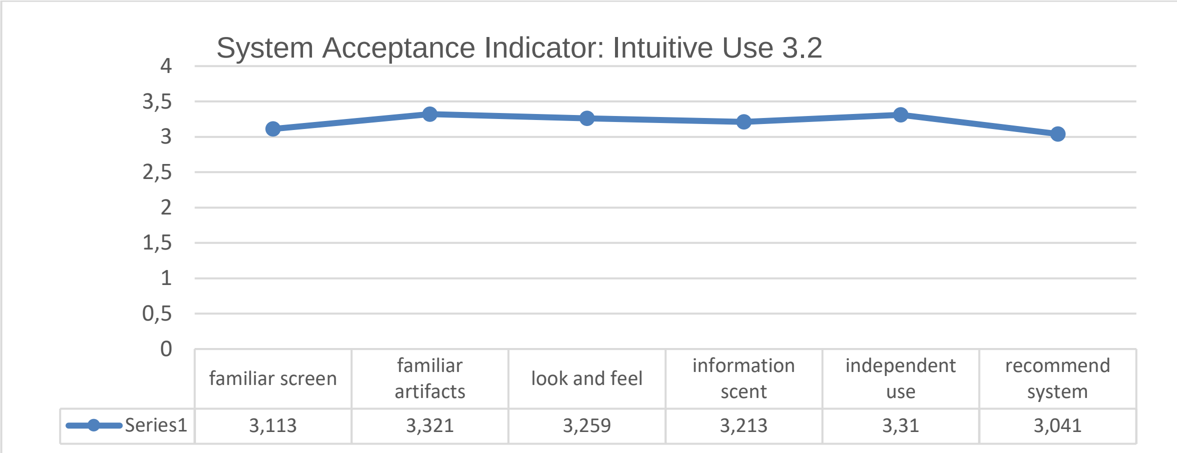


Figure 9. Graph of SAI intuitive use analysis. The Intuitive Use value 3.2 value (overall) results from the average of the responses in the series.

design the CLSE system was deemed successful because the Course Evaluation Survey Office serviced only three help requests.

Three academic staff who had not used the help resources contacted the system administrator for assistance. They were directed to consult the help resources and if they still had problems to make contact again. There was no further contact from these people; apparently, they were successful. Thus, this overall outcome justified the time and resources expended on the development of the help resources. Although contacted for assistance, this was the first time that the system administrator did have to interact with the system itself to service requests for help with the CLSE system.

The anecdotal evidence supported the analytical evidence. On this project, exhaustive programming rigor was maintained in meticulously coding the diverse screen layouts and in analyzing and coding the business processes for the staff UIs. Similarly, instructional design rigor produced best practice adult learning in the development of the help resources. This attention to detail enabled the software to meet the specified requirement to be used without the need for formal training, thus providing an example of best practice in systems interaction design.

Statistical Analysis: Methodology and Findings

The statistical package used was R and the methods used are documented in publications by Everitt (2010), Field, Miles, and Field (2012), and Yong and Pearce (2013). For this quantitative analysis, measures of internal consistency were used to confirm that the set of survey items assembled to measure the same general construct produced similar scores. A high internal consistency indicates that the items contribute to the understanding of the overall construct. All survey item responses were on a 1-to-5 Likert scale. As noted earlier, response cases with missing data were removed from the data set for quantitative analysis.

Test for Pairwise Linear Correlation

Pearson's, Spearmen's, and Kendall's pairwise correlations are measures of the linear correlation between two variables. These correlations were calculated from the survey data. All statistics were relatively consistent for the data set. Kendall's tau was selected for the analysis due to the small data set obtained from convenience sampling; the correlations are presented in Table 6.

A low-to-moderate (0.10–0.51) positive correlation was found between the pairs of the survey items on the use of the new CLSE system. Reflective question Item 5 had the lowest correlations. This can be expected due to variations in personal expectations and experiences and each individual's interpretation of these perceptions.

A significance test performed on the lowest correlation between Item 1 and Item 5 was significant (p <.001). All correlations used the same sample size; consequently, all other pairwise correlations would also be statistically significant.

Test for Cronbach Alpha

The Cronbach alpha test was performed on the data set. The test provides an estimate of how well the items measure the general construct. It is a function of the number of items, the average covariance between item-pairs, and the variance of the total scores. Generally the test result increases with increased intercorrelations between items. Cronbach's alpha of 0.7 to 0.8 is considered satisfactory for basic research. The R statistical software functionality for determining Cronbach's alpha also calculates test values for internal consistency: the interitem, item–total and corrected item–total correlations. The test results are presented in full in Appendix B.

Table 3. Kendall's Pairwise Correlations (n = 54).

SAI Question	SAI Q1	SAI Q2	SAI Q3	SAI Q4	SAI Q5	SAI Q6
The screen for this software looks like other screens I have used.	1					
When I look at the icons and menus, etc., on the screen, I know what to use and how to use it.	0.387	1				
For the things that I use, this software looks and works the same way, every time.	0.293	0.468	1			
There is always enough information on the screen when it is needed.	0.439	0.469	0.404	1		
I think I will be able to use this software without asking for help from the experts who know how to use it.	0.118	0.209	0.472	0.483	1	
I would recommend this software to a colleague or friend.	0.333	0.483	0.492	0.499	0.514	1

Note. A significance test on the lowest correlation between SAI Q1 and SAI Q5 was significant (p < .001). All correlations used the same sample size; consequently, all other pairwise correlations would also be statistically significant.

Cronbach's alpha is an estimate of how well the six items measure the computer experiential knowledge required to use the CLSE system. The Cronbach's alpha test result of 0.8 for the data is within the acceptable range.

Values of the overall alpha are calculated with each item not included in the calculation. In this case, these item values should not be greater than 0.8, which was Cronbach's alpha for the test. An item ranking greater than alpha should be dropped from the analysis to improve reliability. No items were greater than 0.8, indicating all items were retained for analysis.

If an item is less than 0.3, that item should not be retained, to improve reliability. Values above 0.5 indicate a strong internal consistency.

The average interitem correlation was 0.400. Interitem correlations are the average of all relevant pairwise correlations. No items were less than 0.3, indicating that all elements correlate well and are measuring a characteristic of a single construct.

The average corrected item–total correlation was 0.562. Item–total correlations are the correlations between each item and the total score for all responses for each case. Corrected item–total, or item–rest correlation, is the correlation of that item with the scale total if that item is not included in the total. The corrected item–total correlation for each case is greater than 0.3, indicating that all items correlate well with the overall scale.

Cronbach's alpha test results indicate that the internal integrity of the data set cannot be improved for quantitative data analysis by the removal of any items from the data set. Consequently, the survey data analysis results have a satisfactory level of correlation to indicate that the data consistently reports on the general concept under evaluation: intuitive use.

Principal Factor Analysis

Principal factor analysis derives a mathematical model from the correlation matrix. The matrix eigenvalues associated with an item indicate the relative importance of the factors. A scree plot or the eigenvalues can locate the cutoff for selecting factors. Eigenvalues greater than one also help to indicate which factors are to be retained for further analysis. To calculate the degree that the items load on each factor, a technique called rotation is used to discriminate between factors.

Principal factor analysis was undertaken to reduce the data set into a smaller subset of measurement items and to apply that outcome to the findings as a descriptive method for the sample collected. A series of principal factor analyses were run to determine the number of factors for analysis and the interdependence of the factors. The final analysis is presented in Table 7.

The relatively small data set ($n = 54$) was tested for suitability for factor analysis. In Test 1, the Bartlett-Sphericity test, the p-value of 2.22e-16 is less than 106.506, indicating it is unlikely that the observed correlation matrix was derived from a population with zero correlation. The data may provide meaningful analysis results.

Test 2, the Kaiser-Meyer-Olkin (KMO), which included Measures of Sampling Adequacy (MSA), is referred to in Table 8. The $MSA > 0.5$, therefore all elements can be included in the analysis. KMO-Criterion was 0.735. Test 3 was the Mod R test to ensure no items are very highly correlated (multicollinearity) and checks that the determinate of the correlation matrix > 0.0001. The determinate was 0.154. The Bartlett's Test result in conjunction with the KMO test result and Mod R indicate that the data can be used to perform a valid factor analysis.

Table 7. Principal Factor Analysis Results.

SAI Questions	Factor 1	Factor 2	Factor 3
SAI Q2	0.88		
SAI Q3	0.75		
SAI Q5		0.94	
SAI Q4		0.52	0.59
SAI Q6	0.45	0.47	
SAI Q1			0.91
Eigenvalues	1.72	1.64	1.34
Proportion variance	0.29	0.27	0.22
Cumulative variance	0.29	0.56	0.78
Proportion explained	0.37	0.35	0.28
Cumulative proportion	0.37	0.72	1.00

Note. Dominant Factor 1 accounted for 29% of the variance, Factor 2 accounted for 27%, and Factor 3 accounted for 22%. Together they account for 78% of the variance.

Table 8. All Measures of Sampling Adequacy for the SAI Questions Were Greater than 0.5 and Included in the Analysis.

SAI Q1	SAI Q2	SAI Q3	SAI Q4	SAI Q5	SAI Q6
0.7136030	0.7252974	0.780568	0.7630045	0.6260568	0.7839342

The Factors

The interpretation of the factors is dependent on the items that load onto each factor. The questions with the highest loading dictate the character of the factor. The following names of the factors reflect this loading.

Factor 1–Familiar Use Expectations:

- SAI Q2 reflects effort expectancy and an expectation for clear and understandable control for ease of use;
- SAI Q3 reflects effort expectancy and pragmatic assessment of control for ease of use;
- SAI Q6 reflects ownership from positive user experience.

Factor 2–Confident Interaction Experiences:

- SAI Q5 reflects facilitating conditions and an expectation for self-efficacy with respect to control and usefulness;
- SAI Q4 reflects facilitating conditions and pragmatic assessment of control using screen artifacts;
- SAI Q6 reflects ownership from positive user experience.

Factor 3–Recognition of Prior Learning:

- SAI Q1 reflects visceral response to effort expectancy after observing the screen artifacts and layout;

- SAI Q4 reflects facilitating conditions and pragmatic assessment of control using screen artifacts.

Cronbach's alpha test provides a measure of the reliability of the factors for this survey. Factor 1 has Cronbach's alpha = 0.75, Factor 2 = 0.745, and Factor 3 = 0.67. The removal of any items from the factors would not improve their reliability.

The three factors—Familiar Use Expectations, Confident Interaction Experiences, and Recognition of Prior Learning—are all components identified in current research into intuition. Together they describe the internal consistency of the survey questions that characterize the general construct: intuitive design.

The survey data analysis results have a satisfactory level of correlation and factor extraction to indicate that the data consistently reports on the general concepts under evaluation. Consequently, the six items used for this evaluation as a subset of the SAI are considered suitable for evaluating intuitive design and intuitive use of software. Additional research to confirm this preliminary finding is recommended. However, based on the analysis presented, there is a reasonable basis for further use of the question set for the evaluation of software for intuitive design and use.

CONCLUSIONS

The analysis of the survey in the case study resulted in an SAI Intuitive Use evaluation of 3.2. On a scale of 0 to 4, this result is indicative of a perception of a positive intuitive use experience. A factor analysis of the survey responses exposed three factors that described this intuitive use experience. The three factors in order of impact on intuitive use in the case study were called

1. Familiar User Expectations;
2. Confident Interactions;
3. Leverage of Prior Learning.

These factors replicate the theoretical elements of current research into intuition and are a one-on-one mapping of them onto the criteria for intuitive design and subsequent intuitive use. At this early stage, the survey has produced initial results confirming the proposed structure for analysis of intuitive design and intuitive use. This structure can assist further research into intuitive design and use.

The delivery of an experience of intuitive use can be evaluated by a survey of six questions that link the experience to an HCI framework of intuitive design criteria (see Table 9). This question set can be used by other researchers in the validation of their user interfaces and the results of the survey included in stakeholder return on investment reports.

In this paper I discussed current research findings into intuition and intuitive use and applied them to existing HCI design guidelines in a case study. There are two requirements for an experience of intuition: (a) existing experiential knowledge and skills in a particular domain, and (b) an unexplainable perception that a novel situation is contextually familiar. Extending intuition research into intuitive use adds the caveat of (c) successful application of previously acquired experiential knowledge and skills. The survey question set used in the case study assess the successful use caveat.

Table 9. Intuitive Design and Use Survey Questions Mapped to the HCI Design Criteria for Intuitive Interaction.

Questions to assess intuitive use	HCI Design Criteria
The screen for this software looks like other screens I have used.	Usability, distributed cognition, affordances
When I look at the icons and menus, etc., on the screen, I know what to use and how to use it.	Usability, distributed cognition, affordances
For the things that I use, this software looks and works the same way, every time.	Usability, distributed cognition, affordances
There is always enough information on the screen when it is needed.	Usability, distributed cognition, affordances
I think I will be able to use this software without asking for help from the experts who know how to use it.	Depth of immersion in the CoP
I would recommend this software to a colleague or friend.	Affirmation of software ownership and participation in the promotion of the CoP

To facilitate intuitive use, intuitive design has three overriding principles:

- present that which is already known and familiar in the domain;
- use familiar things to explain the use of unfamiliar things in this domain;
- provide consistent internal representation of all that is presented on the screens.

Intuitive use is the observable behaviors that are the direct result of the cognitive process called intuition that accesses previously acquired experiential knowledge and skills to commence and successfully complete a series of actions in a situation not previously encountered. Further research is necessary to confirm these early results.

IMPLICATIONS FOR THEORY AND APPLICATION

This paper reported on action research that implemented current intuitive research tenants to design, implement and analyze a successful enterprise system. Software and in particular interface design is often considered more an art than a science. Current interface design practice, focused on intuitive design as discussed, is often implicit and the expertise built up over time. The research into intuitive design and use is ongoing. This papers contribution is to consolidate design practice and evaluation on a theoretical base that provides both research and industry practioners with an evidence-based theoretical reference point from which to commence their work. The use of the easily applied explicit guidelines for intuitive interface design presented in this paper can result in consistently relevant contextually designed user interfaces. Interfaces designed in this manner can improve the user experience and add to the research into intuitive use. In addition, the six-question survey is a system analysts' tool that can be used to assess any user interface for intuitive use and to confirm the appropriate application of the intuitive design guidelines.

REFERENCES

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist, 33*, 344–358.

Bandura, A. (1986). *Social foundations of thought and action*. Engle Cliffs, NJ, USA: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY, USA: W. H. Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1–26.

Blackler, A. (2006). *Intuitive interaction with complex artifacts* (Unpublished doctoral dissertation). Queensland University of Technology, Australia.

Blackler, A. L., & Hurtienne, J. (2007). *Towards a unified view of intuitive interaction: Definitions, models and tools across the world*. Retrieved from http://eprints.qut.edu.au/19116/

Blackler, A. L., & Popovic, V. (2015). Towards intuitive interaction theory. *Interacting with Computers, 27*(3), 203–209.

Blackler, A. L., Popovic, V., & Mahar, D. P. (2009, September). *Concepts of intuitive interaction.* Paper presented at Workshop Tagung Mensch & Computer, Berlin, Germany.

Bødker, S. (1991). *Through the interface: A human activity approach to user interface design*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Buxton, B. (2010). *Natural user interfaces (NUI)*. Retrieved from http://www.billbuxton.com

Carroll, J. (Ed.). (2003). *HCI models, theories and frameworks: Towards a multi-disciplinary science*. San Francisco, CA, USA: Elsevier Science.

Carroll, J., & Rosson, M. (1987). Paradox of the active user. In J. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interaction* (pp. 80–111). Cambridge, MA, USA: MIT Press.

Checkland, P. (1999). *Systems thinking, systems practice – Includes A 30 Year Retrospective.* Chichester, UK: John Wiley and Sons.

Embrey, D. (2015). *The skill, rule and knowledge based classification* [Monograph]. Retrieved from http://www.humanreliability.com/articles/UnderstandingHumanBehaviourandError.pdf/

Everitt, B. E., & Hothorm, T. (2010). *A handbook of statistical analysis using R* (2nd ed.). Boca Raton, FL, USA: CRC Press.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, UK: SAGE Publications.

Gibson, J. (1979). *The ecological approach to visual perception*. Boston, MA, USA: Houghton Mifflin.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA, USA: MIT Press.

Hutchins, E. (2000, May). Distributed cognition. In *The international encyclopedia of the social and behavioral sciences*. Retrieved from http://eclectic.ss.uci.edu/~drwhite/Anthro179a/DistributedCognition.pdf

Kaptelinin, V. (2014). *Affordances and design.* Cambridge, MA, USA: The Interaction Design Foundation, MIT Press.

Knowles, M. (1984). *The adult learner: A neglected species* (3rd ed.). Houston, TX, USA: Gulf.

Lehane, P. (2008, November). *Usefulness emergent in the user experience: An empirical and theoretical analysis of technology acceptance in the workplace.* Paper presented at International Conference on Kansei Engineering and Affective Systems, Nagaoka, Japan.

Lehane, P. (2010). Designing interactions for the collective user experience. In *Proceedings of the 22nd Conference of the Computer–Human Interaction Special Interest Group of Australia on Computer–Human Interaction.* New York, NY, USA: ACM. https://doi.org/10.1145/1952222.1952247

Lehane, P. (2012a). *The system acceptance indicator: An analysis practice emergent from the plans, scripts and intuitive interaction of the collective user community* (Unpublished doctoral dissertation). University of South Australia, Australia.

Lehane, P. (2012b). Make it intuitive: An evaluation of interaction design using plans and scripts from the computer-community of practice. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *8*(1), 77–103.

Lehane, P. (2012c). Mapping the user experience: Development of a validated instrument from the plans and scripts of the computer community of practice. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments, 8*(2), 158–179.

Lehane, P., & Huf, S. (2007, August). *Re-establishing user perception norms: How long does it really take*. Paper presented at European Conference of Cognitive Ergonomics, London, England.

MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99.

Nardi, B. (Ed.). (1996). *Context and consciousness: Activity theory and human-computer interaction,* Cambridge, MA, USA: MIT Press.

Nielsen, J. (1993). *Usability engineering.* San Diego, CA, USA: Morgan Kaufman.

Norman, D. (1988). *The design of everyday things*. New York, NY, USA: Doubleday.

Norman, D. (2004). *Emotional design: Why we love (or hate) everyday things*. New York, NY, USA: Basic Books.

Norman, D. (2010). Natural user interfaces are not natural. *ACM Interactions*, *17*(3), 6–9.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, *3*(3), 301–314.

O'Brien, M. A., Rogers, W., & Fisk, A. D. (2010). Understanding the role of experience in younger and older adults' interactions with a novel technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54*(20), 1827–1831. https://doi.org/10.1177/154193121005402018

Preece, J., Rogers, Y., & Sharp, H. (2007). *Interaction design: Beyond human–computer interaction* (2nd ed.). New York, NY, USA: Wiley.

Rasmussen, J. (1994). *Cognitive systems engineering.* New York, NY, USA: John Wiley & Sons Inc.

Rogers, A. (2002). *Teaching adults* (3rd ed.). Philadelphia, PA, USA: Open University Press.

Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. In Y. Vassilioun (Ed.), *Human factors in interactive computer systems* (pp. 57–69). Norwood, NJ, USA: Ablex Publishing.

Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: Strategies for effective human–computer interaction* (5th ed.). Reading, MA, USA: Addison-Wesley Publ. Co.

Suchman, L. (1987). *Plans and situated actions*. Cambridge, MA, USA: Cambridge University Press.

Sweller, J. P., Ayres, P., & Kalyuga, S. (2011). *Explorations in the learning sciences, instructional systems and performance technologies, cognitive load theory*. London, UK: Springer Science + Business Media.

Venkatesh, V., Morris, M., Davis, G., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478.

Vicente, K. (1999). *Cognitive work analysis*. Mahwah, NJ, USA: Lawrence Erlbaum Associates Inc.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA, USA: Harvard University Press.

Winston, R. (1970, August). *Managing the development of large software systems*. Paper presented at IEEE WESCON/1970, Los Angeles, CA, USA.

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79–94.

Author's Note

All correspondence should be addressed to
Patrick Lehane
University of Southern Queensland
P. O. Box 32 Darling Heights
Queensland, Australia, 4350
patlehane101@gmail.com

Human Technology: An Interdisciplinary Journal on Humans in ICT Environments
ISSN 1795-6889
www.humantechnology.jyu.fi

APPENDIX A

Descriptors for experience with course-level evaluation assess the familiarity of the academic staff with course-level evaluation per se, any experience with the university's process, and finally if a staff member has developed or researched course-level evaluation.

None: New to academia with no exposure to course-level valuation as someone being evaluated.

Limited: Know of but have no practical experience with course-level evaluation beyond my role as a lecturer.

Moderate: Understand the concepts and methodologies of course-level evaluation, may have carried out own course-level evaluations, have participated in previous course evaluations at the university or another university in my capacity as lecturer or course examiner.

Considerable: Applied analysis derived interventions in the practice of course-level evaluation based on previous course evaluation instruments, may have established course-level evaluation methodologies.

Extensive: Researcher with publications on course-level evaluation.

Descriptors to rate IT skills that relate to basic use and overall familiarity with university supported software that staff are expected to use on a day to day basis.

Beginner: I am learning to use basic programs such as Word and Excel, and I seldom use the Internet or e-mail.

Novice: I use email and the Internet regularly and I have a basic understanding of programs such as Word and Excel.

Proficient: I am an experienced user of email and the Internet and I regularly use programs such as Word, Excel and/or discipline-specific software.

Expert: I can customize an operating system and produce complex documents and presentations in appropriate programs, e.g., Word, Excel, and PowerPoint. Where applicable, I use discipline-specific software to a professional level.

APPENDIX B

Cronbach's alpha test is a measure used to assess the internal consistency, that is, the reliability of a set of test data. It is calculated by correlating the score for each individual survey response with the total score for each observation and then comparing that to the variance for all individual item scores.

The overall α for the test is the raw_alpha value of 0.79 with a mean of 3.9 and a standard deviation of 0.46. In Table B1, each row in the reliability if an item is dropped section refers to an SAI question and its associated raw_alpha. If any of the raw_alpha values are greater than the overall α of 0.79, dropping that particular item will increase the overall reliability of the data set analyzed. The result for X29.1 (SAI Question 1) is in the questionable range but was retained because other test results indicated that dropping it would not improve the reliability of the data set. The other columns of this section of the table refer to how the other statistics will change if that particular question is dropped/deleted.

In the item statistics section, if r.drop values are less than about 0.3, it means that particular item doesn't correlate very well with the scale overall and should be dropped from the data set. X29.1 has an r.drop value of 0.4 indicating it should be retained.

Table B1. Cronbach's Alpha Test for the SAI Intuitive Use Question Set.

```
Reliability analysis
Call: psych::alpha(x = MyOpinionHCI6.noNa, keys = c(1, 1, 1, 1, 1,
    1))

  raw_alpha std.alpha G6(smc) average_r S/N  ase mean   sd
      0.79       0.8    0.81      0.39 3.9 0.04  3.9 0.46

 lower alpha upper     95% confidence boundaries
0.71 0.79 0.87

 Reliability if an item is dropped:
      raw_alpha std.alpha G6(smc) average_r S/N alpha se
X29.1      0.80      0.80    0.79      0.45 4.0    0.040
X29.2      0.75      0.76    0.75      0.38 3.1    0.049
X29.3      0.76      0.76    0.77      0.39 3.2    0.046
X29.4      0.74      0.74    0.75      0.37 2.9    0.052
X29.5      0.78      0.79    0.76      0.42 3.7    0.045
X29.6      0.73      0.73    0.74      0.35 2.7    0.055

 Item statistics
       n raw.r std.r r.cor r.drop mean   sd
X29.1 57  0.59  0.58  0.46   0.40  3.8 0.67
X29.2 57  0.72  0.73  0.67   0.60  4.0 0.55
X29.3 57  0.68  0.71  0.62   0.54  4.0 0.58
X29.4 57  0.78  0.77  0.72   0.64  3.9 0.72
X29.5 57  0.61  0.64  0.56   0.48  4.2 0.50
X29.6 57  0.82  0.80  0.77   0.67  3.7 0.86

Non missing response frequency for each item
         2    3    4    5 miss
X29.1 0.05 0.19 0.67 0.09    0
X29.2 0.00 0.16 0.70 0.14    0
X29.3 0.02 0.12 0.72 0.14    0
X29.4 0.05 0.18 0.63 0.14    0
X29.5 0.00 0.05 0.72 0.23    0
X29.6 0.07 0.33 0.40 0.19    0
```

Human Technology

www.humantechnology.jyu.fi

humantechnology@jyu.fi

Submissions: humantechnologypublishing.jyu.fi