SUBMISSION | DIETER W. JAHNS STUDENT PRACTITIONER AWARD 2020

CASSANDRA: A decision support tool for **c**linical **ass**essment **and** **r**easoning in **a**nesthesia

Sara Wolf (Main author) & Franzisca Maas (Secondary author + team member)

Students of „Human-Computer Interaction“

Institute Human-Computer-Media, Julius-Maximilians-Universität Würzburg, Germany
in cooperation with
Department of Anesthesia and Critical Care, University Hospital Würzburg, Germany

Project supervisor: Dr. Tobias Grundgeiger

Project realised: September 2018 – August 2019

Julius-Maximilians-UNIVERSITÄT WÜRZBURG

Universitätsklinikum Würzburg

1. STUDY OBJECTIVE

A general task of anesthesiologists is to induce and maintain general anesthesia into a patient. This task includes to rapidly, safely, and pleasantly produce and maintain amnesia, analgesia, akinesia, and an automatic and sensory block while maintaining hemodynamic stability and sufficient ventilation (King & Weaving, 2017). Oftentimes, the anesthesiologists' task is described by the term "hours of boredom and seconds of terror" (Gaba, Fish, Howard, & Rall, 1998). Especially in the "seconds of terror", the operating room (OR) crises, anesthesiologists are required to make decision regardless of whether all information is present and knowing that inadequate decisions may have fatal consequences (Stiegler & Ruskin, 2012).

Therefore, a range of analogue and digital support tools have been developed that aim at influencing the decision-making (Berner & La Lande, 2007). While previous support tools such as checklists have been developed to support the treatment of specific (but rare) diagnoses such as Malignant Hyperthemia (Runciman et al., 2005), support on how to decide on a specific diagnosis in the "seconds of terror" is infrequently addressed. In addition, many support tools in anesthesia as well as in other clinical contexts have failed when being implemented in the actual clinical context (e.g., Elwyn et al., 2013; Jaspers, Smeulers, Vermeulen, & Peute, 2011; Kawamoto, Houlihan, Balas, & Lobach, 2005). Research suggests that the minimal consideration of the actual clinical context during design and evaluation is a potential cause (Musen, Middleton, & Greenes, 2014; Wears & Berg, 2005). A review of the current human-computer-interaction (HCI) literature on support tools for the clinical context supports this issue: The contextual fit of decision support tools (DSTs) has rarely been addressed and the few examples give insights on an exploratory basis without answering why a specific design improved or worsened the contextual fit (e.g. (Yang, Steinfeld, & Zimmerman, 2019; Yang, Zimmerman, Steinfeld, Carey, & Antaki, 2016)). We identified theories of embodied contextual fits (Dourish, 2004; Grundgeiger, Hurtienne, & Happel, 2020; Van Dijk & Hummels, 2017) and needs-based (Grundgeiger et al., 2020; Hassenzahl, Diefenbach, & Göritz, 2010) user experience (UX) to potentially explain the phenomena of (successful and unsuccessful) contextual fits.

Given the lack of proper support tools for the diagnostic process and treatment in the "seconds of terror" and the rarely considered context, our objective was (1) to understand the complex, safety-critical context of OR crises based on UX theories, (2) to develop a decision support tool (called Cassandra) for diagnosis and treatment guided by UX theories, and (3) to evaluate the developed solution and the theoretical framing in a simulator-based study. To do so, we followed a user-centered design process as shown in Figure 1 that addressed all ergonomics core competencies. The present submission is a summary of the overall project and focusses on the development and preliminary evaluation of Cassandra. For a more detailed analysis of the theoretical framing, please refer to our paper accepted at the ACM conference on Designing Interactive Systems (DIS) 2020 ((Klüber et al., 2020), see Appendix A). Text sections from the publication are presented in this submission.

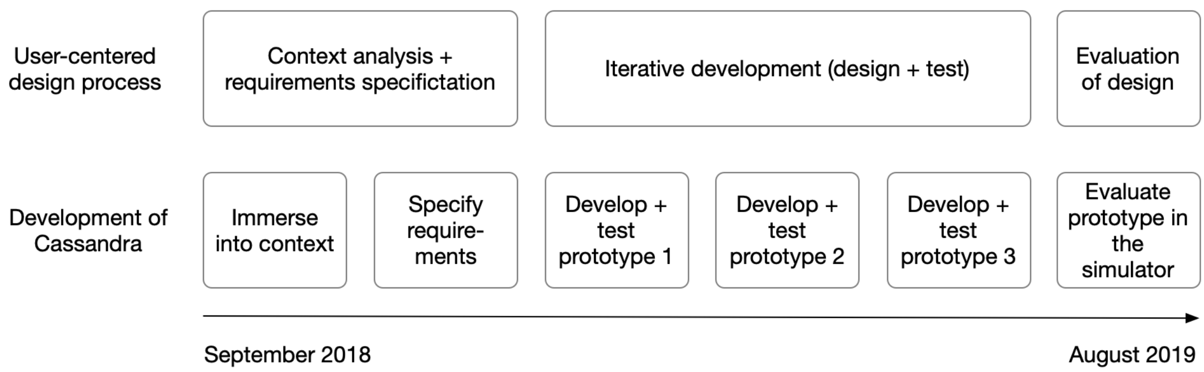

Figure 1. Overview of the development process of Cassandra along the user-centered design process.

2. CONTEXT ANALYSIS AND REQUIREMENTS SPECIFICATION

The overall project was conducted with the anesthesia department of the University Hospital of Würzburg. In a first step, we met with anesthesiologists (1 supervising and 1 trained junior anesthesiologist) from the University Hospital of Würzburg to get an impression of their work and their daily concerns. Inspired by this meeting and the subsequent literature review on the current state of the art, we specified the context to OR crises in anesthesia that was particularly relevant to the collaborating anesthesiologists and underexplored in the related literature.

Commonly, one anesthesiologist and one anesthetic nurse accompany each surgery. In addition, they may call a supervising anesthesiologist for support. During OR crises, anesthesiologists have to diagnose and treat a patient within minimal time in order to keep the patient stable. The overall process involves the whole anesthetic team that tries to perceive, understand, and relate different symptoms, data, and information to identify a coherent diagnosis and to subsequently treat the patient accordingly. Therefore, a DST needs to support both diagnosis and treatment within seconds and a good contextual fit is vital when a DST is ought to be used during OR crises.

To better understand the dynamics of the OR crises, we followed the contextual design approach (Holtzblatt & Beyer, 2017). Given the fortunately infrequent emergencies, contextual inquiries were not possible. Instead, we conducted four semi-structured interviews with anesthesiologists of varying experience levels (supervising, five years, one year). Additionally, we inspected two videos of simulation scenarios and were present at three simulator training scenarios (twelve medical students in their last year). All participants were recruited from the University Hospital of Würzburg. We consolidated the data by organizing the notes into an affinity diagram. The diagram was used to formulate key insights (requirements) and generate visions that would inspire prototype designs.

Overall, we identified three groups of key insights. The first group related to the anesthesiologists' *need for autonomy and competence* and included that the system should not patronize its users, respect different levels of experiences, be trustable and acceptable for its users, and not make users feel guilty in any way. In other words, the system should not make users feel that they should have already known displayed information and that the system displayed the information just because the users' performance was not "good" enough. The second group related to the *environmental context* and included that the system should support clinicians in both calm and stressful situations, only require minimal interaction to avoid additional stress for clinicians, and be able to handle different kinds of data (e.g., continuous vital sign parameters or specific patient information). The third group concerned *teamwork and communication* and included that the system should enable all team members to have the same information about the patient and be accessible for the whole team, and support the joint process of decision-making.

We developed 22 visions concerning teamwork, display concept, and deciding on a diagnosis. We evaluated all visions along the key insights. Finally, the best matching ideas were merged in an overall concept that was then realized in a first paper-based prototype.

3. ITERATIVE DEVELOPMENT: DESIGN + TEST OF CASSANDRA

Overall, we ran three iterations of user tests and interviews with different versions of the Cassandra prototype to validate and improve the design. In each iteration, an interviewer showed the prototype to one user at a time, and another researcher took notes. Users were asked to think of an imaginary scenario of a standard anesthesia procedure or intraoperative crisis they recently experienced during work and browse through the prototype commenting on the interaction and what should be improved. Overall, one human factors specialist, six senior anesthesiologists (work experience >5 years each), two junior anesthesiologists (work experience <5 years each), and one anesthetic nurse gave feedback on different versions of the prototype. The first version was a paper prototype, and versions two and three were digital wireframe prototypes presented on a laptop. After each session with users, we consolidated the issues and emerging requirements, which were then implemented before the next iteration.

While the present submission especially focusses on the development and interaction of the digital tool Cassandra, the content for Cassandra was also generated. Briefly summarizing this process, a group of three anesthesiologists developed and iterated on a set of symptoms and their relations to 15 differential diagnoses of malignant hyperthermia (MH) to establish the content for a test scenario of Cassandra (as necessary for the evaluation). All diagnoses shared symptoms with MH, which is a rare, potentially life-threatening incident with non-specific early symptoms that could easily be confused with many differential diagnoses (Schneiderbanger, Johannsen, Roewer, & Schuster, 2014).

Figure 2. Staged photo of an anesthetic team interacting with Cassandra in the simulated OR environment.

The final concept of Cassandra is a DST that can be accessed via a touch screen, which is positioned at the drape separating surgeons from anesthesiologists during surgery (see Figure 2). Cassandra has three major functions. (1) While it is not activated, the center of the screen shows a video stream of the surgical side of the OR to support situation awareness, communication, and teamwork between the anesthetic and surgical team. In addition, the patient's current vital signs are always displayed on the right side so that the anesthetist does not need to turn away from the patient anymore as is currently typical. To satisfy the anesthesiologists' need for autonomy, we decided to only display the (2) diagnosis feature of Cassandra on demand (see Figure 3). The diagnosis function is activated by entering a potential diagnosis using speech recognition or a keyboard. The searched diagnosis then appears in the middle of the screen accompanied by possible differential diagnoses beneath it. Cassandra visualizes the different symptoms and signs a diagnosis "consists of". Each of these signs can be in one of three states represented by color and placement in relation to the diagnosis container: "present" (green; within), "undetectable" by Cassandra (blue; on the edge), or "absent" (grey; outside). For each diagnosis, the signs are arranged based on their importance from left to right. Cassandra automatically judges and updates each sign based on all digitally available data (demand for minimal interaction) and shows a single check mark that indicates a system-side check. The diagnosis display can be used as a cognitive aid for which symptoms to check in which order to verify a diagnosis. The clinician may change the signs' state by swiping them in or out of the diagnosis container according to his or her assessment (need for autonomy and competence). States entered or confirmed by the anesthesiologist will be marked by two check marks. The diagnosis display enables a rapid evaluation of whether the assumed diagnosis is supported (many signs are green) or whether another differential diagnosis is more likely. The alternative diagnoses are sorted by probability and adjusted at the bottom half of the screen whenever states change. These diagnoses are displayed using the same visual representation as the main diagnosis, however, smaller and without inscriptions. Displaying further differential diagnoses has several advantages: Even if the anesthesiologist searched for a "wrong" diagnosis, alternative diagnoses are visible at a glance and the anesthesiologist may decide to consider a more probable diagnosis or consciously discard an alternative. Moreover, by being confronted with alternative diagnoses and signs, typical cognitive errors/biases such as fixation errors and confirmation bias may be prevented. The set of further diagnoses may also stimulate reflection on different possibilities within the team. Finally, the visual presentation of Cassandra supports staff that have been called for help in perceiving and understand the current situation. Cassandra received its name from the anesthesiologists after they explored the above described functionality. The name is an acronym for "**c**linical **ass**essment **and r**easoning in **a**nesthesia" and a reference to Greek mythology: If people had believed in Cassandra's ability to predict the future, great damage could have been prevented.

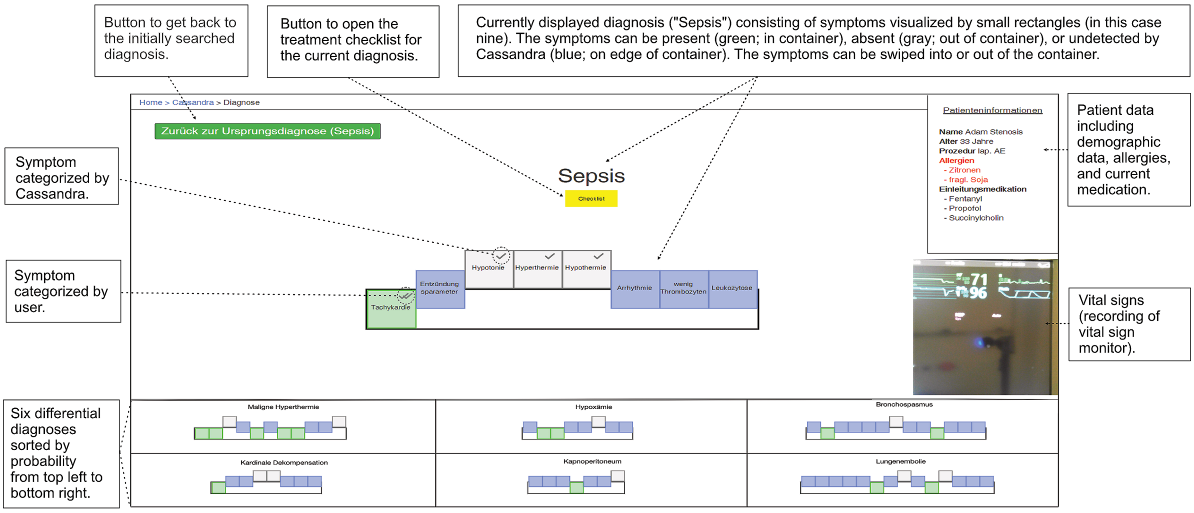

Figure 3. Screenshot of the final Cassandra prototype with explanations. Cassandra shows the diagnosis "Sepsis" and the most likely differential diagnoses for a simulated scenario.

When the anesthesiologist decided on a diagnosis, (3) steps for treatment can be accessed. The treatment information within Cassandra based on the aforementioned checklists (but in an easily accessible format) can be retrieved. In addition, checked off tasks are visible for the whole team and can support better allocation of human resources and teamwork.

Having finalized the design together with the anesthesiologists, the next step was to validate the overall concept within the context: an OR crisis. At this stage of development, we decided to not install the prototype in a real OR but rather run a high-fidelity, full-scale patient simulator study (see Figure 2). Such setting closely resembles realistic circumstances of an ongoing procedure in the OR and is, therefore, an appropriate method to evaluate new technology (Cumin, Weller, Henderson, & Merry, 2010).

4. EVALUATE DESIGN

4.1 Method

Participants. In total, nine participants (four anesthetic nurses, five anesthesiologists) were recruited from the University Hospital of Würzburg. Participants took part in teams consisting of one nurse and one anesthesiologist (and in one case two anesthesiologists). The average age of the participants was 31 years (SD = 6.53) and their working experience varied between 0.5 and 13 years (*M* = 6.2, *SD* = 4.8). Participants received a five Euro cafeteria voucher. All participants had taken part in simulation-based training before. None of the participants had experienced a case of malignant hyperthermia in real life.

Materials and Procedure. Participants arrived, were informed, gave consent, and received a short, standardized introduction to familiarize them with Cassandra. Next, the participants left the simulation room, and the simulation started. Each scenario included two confederates acting as surgeons who simulated a laparoscopic appendectomy in a high-fidelity patient simulator (HPS Human Patient Simulator; CAE Healthcare, Sarasota, FL, USA) and one confederate anesthesiologist. The participants were introduced into the scenario via a handover from the confederate anesthesiologist who then left the simulation room and did not re-enter the scenario. After the handover, the previously stable patient soon started to deteriorate, and participants were required to provide diagnosis and treatment. All scenarios were audio-video recorded using two cameras (GoPro HERO4®).

After the scenario, we conducted a semi-structured interview that aimed to understand the participants' experience in detail, asking questions like *"How did the system fit into your workflow and your thought processes during the operating room crisis".* Although we prepared an initial set of questions, we were interested in any aspect that the participants raised during the interview. All interviews were audio-recorded, and the overall session took around 45 to 60 minutes.

Cassandra was presented on an Iiyama T2735 touchscreen (13.2" x 23.5"), which was positioned in front of the green drape (see Figure 1). During each scenario, two anesthesiologists controlled the patient simulator from an adjoining control room. A researcher in the control room controlled Cassandra in a wizard-of-oz manner. The researcher imitated the automatic detection of changing symptoms according to the state of the patient simulator. To realize the "window" functionality and to provide the current vital sign parameters, the patient monitor and a section of the surgical side were recorded with two webcams (Logitech HD Pro Webcam C920) and displayed in real-time at the intended positions within Cassandra.

Evaluation Procedure. To gain insights into the contextual fit, we analysed both scenario and interview recordings. The videos of each scenario were reviewed independently by two researchers, and sequence models of all interactions with the system were developed and consolidated across all participants (Holtzblatt & Beyer, 2017). The interview data was thematically coded following the content-structuring qualitative analysis as described by Mayring (2008). In this text, we will only briefly summarize our main findings regarding the design of DSTs, however a more in-depth analysis of the evaluation can be found in Klüber et al. (2020) (Appendix A).

4.2 Results and Discussion

We structure this section along some suggestions for DST designs that we will explicate and discuss with examples from our results.

Leave decisions to the user (autonomy). In order to design for autonomy, we decided to leave decisions to the user in several different ways. For example, Cassandra did not initiate an interaction but had to be called by the anesthesiologists. During the evaluation however, most participants needed a hint (delivered by an instructed surgeon) to start interacting with Cassandra during the OR crisis. It is to be seen in future experiments whether this specific design for autonomy (in combination with users possessing little experience with the tool and an operating room crisis) might have led to non-use and should, therefore, be avoided, or whether this handling is vital for the general acceptance of Cassandra and should be preserved. The results from the interviews suggest that leaving decisions to the user is vital for a DST's acceptance. For example, one participant stated that in their opinion, *"the decision-making process can´t be taken away from the anesthesiologists. The system narrows down possibilities, and maybe you discover something that you didn´t consider before, but the decision in which direction you want to go is not taken away by the system. I wouldn´t like it if it did [take the decision away]"* (A3). The statement indicates that our system supported rather than patronized clinicians during usage. The system kept the participants' level of self-involvement high.

Provide relevant information suitable to the workflow to support decision-making (coupling with the social + physical environment) and to subsequently make them feel they are doing better work (competence). Cassandra's capacity to fit into the anesthesiologists' workflow and way of thinking and to satisfy their need for competence, even in anesthesiologists with different levels of experiences, became clear within the video analysis. Experienced anesthesiologists entered the most probable diagnosis as the first tentative diagnosis, while less experienced anesthesiologists did not. Nonetheless in both cases, the anesthesiologist was able to identify the most probable diagnosis by being confronted with several differential diagnoses compared to the diagnosis entered (see Figure 2, in which the first differential diagnosis is most probable). While entering a tentative diagnosis allows anesthesiologists to stick with their standard way of thinking during an operating room crisis (type 1 thinking: quick, intuitive, error-prone (Pelaccia, Tardif, Triby, & Charlin, 2011)), Cassandra uses this input to display structured information on differential diagnoses (type 2 thinking: slow, analytical (Pelaccia et al., 2011)). Therefore, both ways of thinking become seamlessly interwoven through the interaction between anesthesiologists and Cassandra, which subsequently may prevent typical cognitive biases such as premature closure (accepting one diagnosis instead of considering reasonable alternatives) or omission bias (the tendency towards inaction rather than action). The above aspect eventually supports the anesthesiologists' competence by strengthening them against common cognitive biases in anesthesia (Stiegler, Neelankavil, Canales, & Dhillon, 2011) and thereby enhancing their performance.

During analysis, we found another example that demonstrated Cassandra's successful capacity to satisfy needs for autonomy and competence. Prompted by blue symptoms, which could only be assessed with the results of a blood gas analysis, all teams requested a blood gas analysis and subsequently reevaluated the diagnosis by reiterating through the symptoms. This example demonstrates that Cassandra does not make any decisions, but only provides information. It allows the users to decide on which actions to perform and thus leaves the perceived locus of competence with the anesthetic team.

Trigger and structure team communication (relatedness and social coupling). In the video analysis, we found that during the reevaluation of the diagnosis upon reception of the blood gas analysis, teams distributed their tasks so that, for example, the nurse would examine the symptoms' state in Cassandra while the anesthesiologist would read the results of the blood gas analysis out loud. This impression of well-structured team communication was confirmed during the interviews. For example, one participant said that the communication *"is better, for sure. [Cassandra helped] to find a joint we-know"* (A3), while another participant stated that they thought, *"that you speak out loud your thoughts more frequently"* (A6). Another participant said: *"you can use it as a communication tool that everyone knows what is going on […] I am currently missing such a possibility, to be honest"* (N7). We interpret this quote as a satisfaction of the need for relatedness. Participants experienced a sense of belonging and closeness that was induced by increased communication behavior. The above quotes also support the idea that the system increased social coupling – fluent interaction between the users. Whether Cassandra led to increased communication among the members of the anesthetic team is an interesting aspect that should be explored in a comparative follow-up experiment.

Limitations. The present study has several limitations. First, although small sample sizes are common in contextual design (Holtzblatt & Beyer, 2017), our sample size in both qualitative explorations and especially the pilot evaluation was rather small. Second, the pilot evaluation focused on qualitative feedback, and we did not evaluate a UX-driven design vs. a non-UX-driven design. Future research should use such an experimental comparison and use quantitative measures concerning psychological need satisfaction (e.g., Hassenzahl et al., 2010). Third, the pilot evaluation was conducted within a simulated medical environment instead of real OR crisis. Fourth, we only considered a set of 15 diagnoses related to the scenario in the pilot evaluation, and future research is needed to investigate the scalability of Cassandra to more than one scenario. Fifth, besides UX, there are certainly other approaches to increase the environmental fit and acceptance of DSTs. Previous work on automation can provide valuable input here (Lee, Wickens, Liu, & Boyle, 2017) as, for example, automation exposure has been found to increase the acceptance of automated processes (Bekier & Molesworth, 2017).

5. Conclusion

To the best of our knowledge, we were the first to develop a DST that takes into account the diagnostic process and the treatment during OR crises. In addition, our design was based on an analysis of the (greater) context involving the physical and social environment as well as human beings' constraints and givens (such as psychological needs or cognitive limitations), which was seldomly considered in prior work. The performed evaluation indicated that our focus on the greater context through the lens of theories of embodied and needs-based UX could be valuable for DST acceptance and contextual fit. However, this remains to be strengthened by future work.

For future DST designs, we recommend taking psychological needs such as autonomy, competence and relatedness into account. For example, we suggest leaving decisions to the user and rather providing relevant information suitable to current workflows. This also contributes to a good coupling with the social and physical environment that is vital for DST designs if they are to work in the context.

References

Bekier, M., & Molesworth, B. R. (2017). Altering user' acceptance of automation through prior automation exposure. *Ergonomics, 60*(6), 745-753. doi:https://doi.org/10.1080/00140139.2016.1216610

Berner, E. S., & La Lande, T. J. (2007). Overview of Clinical Decision Support Systems. In E. S. Berner (Ed.), *Clinical Decision Support Systems: Theory and Practice* (pp. 3-22). Cham, Germany: Springer.

Cumin, D., Weller, J. M., Henderson, K., & Merry, A. F. (2010). Standards for simulation in anaesthesia: creating confidence in the tools. *British journal of anaesthesia, 105*(1), 45-51. doi:https://doi.org/10.1093/bja/aeq095

Dourish, P. (2004). *Where the action is: the foundations of embodied interaction*. Cambridge, MA, USA: MIT press.

Elwyn, G., Scholl, I., Tietbohl, C., Mann, M., Edwards, A., Clay, C., . . . Frosch, D. L. (2013). "Many miles to go…": a systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC medical informatics and decision making, 13*. doi:https://doi.org/10.1186/1472-6947-13-S2-S14

Gaba, D. M., Fish, K. J., Howard, S. K., & Rall, M. (1998). *Zwischenfälle in der Anästhesie: Prävention und Management*: Fischer.

Grundgeiger, T., Hurtienne, J., & Happel, O. (2020). Why and How to Approach User Experience in Safety-Critical Domains: The Example of Healthcare. *Human Factors*. doi:https://doi.org/10.1177/0018720819887575

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with computers, 22*(5), 353-362. doi:https://doi.org/10.1016/j.intcom.2010.04.002

Holtzblatt, K., & Beyer, H. (2017). *Contextual Design: Design for Life*. Burlington, MA, USA: Morgan Kaufmann.

Jaspers, M. W., Smeulers, M., Vermeulen, H., & Peute, L. W. (2011). Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *Journal of the American Medical Informatics Association, 18*(3), 327-334. doi:https://doi.org/10.1136/amiajnl-2011-000094

Kawamoto, K., Houlihan, C. A., Balas, A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj, 330*, 765. doi:https://doi.org/10.1136/bmj.38398.500764.8F

King, A., & Weaving, L. M. (2017). General anesthesia: Induction. Retrieved from https://www.uptodate.com/contents/general-anesthesia-induction?source=search_result&search=general%20anesthesia&selectedTitle=3~150

Klüber, S., Maas, F., Schraudt, D., Hermann, G., Happel, O., & Grundgeiger, T. (2020). *Experience Matters: Design and Evaluation of an Anesthesia Support Tool Guided by User Experience Theory.* Paper accepted at the conference on designing interactive systems 2020 (DIS'20), Eindhoven, Netherlands.

Lee, J. D., Wickens, C. D., Liu, Y., & Boyle, L. N. (2017). Human Automation Interaction. In *Designing for people: An introduction to human factors engineering* (3rd edition ed., pp. 357-387). Scotts Valley, CA, USA: CreateSpace.

Mayring, P. (2008). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Beltz.

Musen, M. A., Middleton, B., & Greenes, R. A. (2014). Clinical decision-support systems. In E. Shortliffe & J. Ciimino (Eds.), *Biomedical informatics* (pp. 643-674). London, UK: Springer.

Pelaccia, T., Tardif, J., Triby, E., & Charlin, B. (2011). An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Medical education online, 16*(1). doi:https://doi.org/10.3402/meo.v16i0.5890

Runciman, W., Kluger, M., Morris, R., Paix, A., Watterson, L., & Webb, R. (2005). Crisis management during anaesthesia: the development of an anaesthetic crisis management manual. *BMJ Quality & Safety, 14*(3), e1-e1.

Schneiderbanger, D., Johannsen, S., Roewer, N., & Schuster, F. (2014). Management of malignant hyperthermia: diagnosis and treatment. *Therapeutics and clinical risk management, 10*, 355-362. doi:http://doi.org/10.2147/TCRM.S47632

Stiegler, M., Neelankavil, J. P., Canales, C., & Dhillon, A. (2011). Cognitive errors detected in anaesthesiology: a literature review and pilot study. *British journal of anaesthesia, 108*(2), 229-235. doi:https://doi.org/10.1093/bja/aer387

Stiegler, M., & Ruskin, K. J. (2012). Decision-making and safety in anesthesiology. *Current Opinion in Anesthesiology, 25*(6), 724-729. doi:http://doi.org/10.1097/ACO.0b013e328359307a

Van Dijk, J., & Hummels, C. (2017). *Designing for embodied being-in-the-world: Two cases, seven principles and one framework.* Paper presented at the eleventh international conference on tangible, embedded, and embodied interaction.

Wears, R. L., & Berg, M. (2005). Computer technology and clinical work: still waiting for Godot. *Jama, 293*(10), 1261-1263. doi:http://doi.org/10.1001/jama.293.10.1261

Yang, Q., Steinfeld, A., & Zimmerman, J. (2019). *Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes.* Paper presented at the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk.

Yang, Q., Zimmerman, J., Steinfeld, A., Carey, L., & Antaki, J. F. (2016). *Investigating the heart pump implant decision process: opportunities for decision support tools to help.* Paper presented at the 2016 CHI Conference on Human Factors in Computing Systems.

A. Appendix

Experience Matters: Design and Evaluation of an Anesthesia Support Tool Guided by User Experience Theory

Sara Klüber
Julius-Maximilians University
Würzburg, Germany
sara.klueber@uni-wuerzburg.de

Franzisca Maas
Julius-Maximilians University
Würzburg, Germany
franzisca.maas@uni-wuerzburg.de

David Schraudt
Julius-Maximilians University
Würzburg, Germany
david.schraudt@protonmail.com

Gina Hermann
Julius-Maximilians University
Würzburg, Germany
gina.hermann@gmx.de

Oliver Happel
University Hospital of Würzburg
Würzburg, Germany
happel_o@ukw.de

Tobias Grundgeiger
Julius-Maximilians University
Würzburg, Germany
tobias.grundgeiger@uni-wuerzburg.de

ABSTRACT

Despite many advances, clinical decision support tools (DSTs) often suffer from implementation and acceptance problems in the actual clinical context. We suggest that considering psychological needs-based and embodied user experience theories in the design of DSTs could help to overcome these problems. To examine this idea, we iteratively developed a DST called Cassandra supporting anesthetic teams in crisis management, specifically focusing on psychological needs and fluent interaction with the social and physical environment. We preliminarily evaluated Cassandra in a medical simulation, requiring anesthetic teams to handle a crisis. Although not all features of Cassandra had the intended effect, the results indicated that interacting with Cassandra supported the fulfillment of the identified needs for autonomy, competence, and relatedness and was seamlessly integrated into existing diagnostic processes. Considering user experience theories for the design of DSTs seems a promising way to overcome implementation and acceptance problems and eventually improve patient safety.

Author Keywords

Decision Support Systems; Healthcare; User Experience

CSS Concepts

• **Information systems~Information systems applications~Decision support systems** • **Human-centered computing~Human computer interaction (HCI) ~HCI design and evaluation methods** • Applied computing~Life and medical sciences~Health care information systems

DIS '20, July 6–10, 2020, Eindhoven, Netherlands.

ACM ISBN 978-1-4503-6974-9/20/07...$15.00.
DOI: http://dx.doi.org/10.1145/ /3357236.3395552

INTRODUCTION

Clinical decision support systems (DSTs) are computer-based systems that aim to influence the decision-making of clinicians [5]. In the context of anesthesia, decision support frequently relates to the diagnosis of a patient's status or a specific treatment [22]. However, in anesthesia as well as in other clinical contexts, many DSTs have failed to be implemented or used in the clinical environment [e.g., 12, 25, 27, 40]. Research has suggested that minimal consideration of the actual clinical context contributed to these failures [35, 50, 53].

In the present paper, we suggest a theory-driven design based on user experience (UX) theories in order to provide the means to fit a DST into the respective clinical context. We make a (1) theoretical, (2) design, and (3) empirical contribution. First, we set out to specifically focus on the clinicians' UX that emerges when interacting with a DST to improve the acceptance of DSTs. In short, we suggest considering psychological needs-based UX and embodied theories of HCI in the design of DSTs. Second, we report a user-centered design process with an emphasis on the above theories that resulted in a DST named Cassandra (**C**linical **ass**essment **and** **r**easoning in **a**nesthesia), designed to support anesthetic teams in the diagnostic process during operating room crises. Third, we report a pilot evaluation of Cassandra that was conducted using a full-scale high-fidelity medical simulation including anesthetic teams (anesthesiologist and anesthetic nurse), multiple actors, a high-fidelity patient simulator, and Cassandra implemented as a wizard-of-oz interface. The results of the pilot evaluation provide a detailed description of how clinicians interacted with the DST and a first evaluation of the suggested theoretical explanation.

RELATED WORK

In this section, we will first describe the diagnostic process in anesthesia and why it is at risk of bias before summarizing DSTs for anesthesia that may resolve bias issues. Finally, we will highlight how UX can provide the means to improve acceptance and environmental fit within the context of DSTs and will introduce UX theory.

Diagnostic Process and Biases in Anesthesia

During an operating room crisis, the anesthetic team is required to rapidly decide on a diagnosis in order to identify the respective treatment. Pelaccia et al. [38] applied the dual-process theory to this diagnostic process: Type 1 processes (intuitive) are quick and automatic processes that are based on prior experiences and therefore may be biased, while type 2 processes (reflective) are slow and analytical. Put in the context of anesthesia, type 1 processes allow anesthesiologists to quickly make decisions in time-critical, complex situations, while type 2 processes might help in identifying solutions to rare issues. Type 1 processes are useful in operating room crises due to their fast processing and the assumption that "common diseases are common" [2]. However, they may also lead to cognitive biases [40].

Cognitive biases are the most frequent cause of diagnostic errors [17] and may result in the wrong treatments and threaten patient safety [8, 45, 46]. In a review, Stiegler et al. [45] presented a cognitive bias catalog specific to anesthesia. For example, premature closure (accepting one diagnosis instead of considering reasonable alternatives) was identified as the single most common cognitive bias leading to diagnostic errors [17].

Support Tools in Anesthesia

To support clinicians in making evidence-based decisions, and prevent cognitive biases, support tools that provide filtered data or knowledge were built [54]. Anesthetic decisions that may be supported are diagnoses, treatments, or prognostic predictions, and the filtered data or knowledge may, for example, be knowledge on differential diagnoses. Most previous support tools focused on providing predictions for more extended time frames (i.e., hours or days) and one-purpose information (e.g., solely treatment such as implement/give medication or not) [22, 36, 44]. During an operating room crisis, however, the anesthesiologist is required to make decisions on diagnosis and treatment within seconds or minutes [46].

While support for treatment during such crises has received considerable attention in the form of support tools such as cognitive aids (artifacts such as checklists that support a user in completing a task), the preceding diagnostic process is seldom explicitly considered or supported by cognitive aids [23, 31]. In addition to several paper-based cognitive aids (e.g., the Stanford manual [14]), researchers also investigated digital aids [e.g., 16, 19, 29, 39, 51]. Nonetheless, cognitive aids (for operating room crises in anesthesia), as well as DSTs, often remain unused or fail when implemented in clinical contexts [31].

A poor contextual fit of support tools in the actual clinical context has been identified as a significant reason for these failures in the cognitive aid [13, 15, 31, 37] and DST [12, 25, 27, 40] literature. Contextual factors include, for example, internal factors of the clinician (e.g., the clinician makes biased decisions due to human cognitive biases, the clinician has specific workflows and routines, the clinician has psychological needs) and the clinician as part of a working team (e.g., the clinician relies on communication with others for decision-making). External factors may also lead to a poor contextual fit (e.g., the clinician being constrained by time-related stress, lack of information, or the actual physical context). Summarized, contextual factors seem to be vital for DST acceptance and need to be considered in DST design.

UX Theory and Support Tools

While most related HCI research on DSTs aims to, for example, improve the presentation and visualization of information [7] or to advance technology [4], initial investigations in advancing the contextual fit were recently made by Yang et al. [53]. Such contributions may be further expanded by taking a closer look into *why* their findings helped to improve the acceptance of the DST. We suggest that psychological needs-based UX and embodied theories of HCI can provide an answer.

UX has been neglected in the context of safety-critical domains [but see 33, 41] and even deemed not relevant [18, 34]. However, whether or not designers pay attention to UX, anesthesiologists always experience an interaction with technology and therefore have a feeling of whether an interaction was "good" or "bad". We will introduce a view on UX based on psychological needs-based and embodied cognition [20].

Psychological needs-based UX theories suggest that good UX emerges when psychological needs, such as autonomy, competence, stimulation, or relatedness, are fulfilled by interacting with a product or service [21]. Neglecting such needs might contribute to non-acceptance, non-use, feelings of discomfort, or even reduced safety. Therefore, considering UX in safety-critical systems seems to make sense [20].

First, the need for autonomy refers to the quality of self-involvement in current activities [43]. Valenta et al. [47] reported that physicians strongly agreed with statements such as *"Alerts intrude on my professional autonomy"* and concluded that support tools could question physicians' competence and intrude upon professional autonomy [see also 51]. This issue matches the psychological need for autonomy. Second, the need for competence refers to attaining or exceeding a standard in one's performance [43]. Concerning DSTs, there is a subliminal fear that technology might replace clinicians' expertise [6, 47, 52]. Furthermore, Wang et al. [49] reported that clinicians explicitly did not want DST information for their initial diagnosis. Both examples indicate that clinicians are concerned that the perceived locus of competence (i.e., the source of performance) shifts towards the DST, and therefore

Figure 1. Staged photo of an anesthetic team interacting with Cassandra in the simulated operating room environment.

compromises need satisfaction. Third, Yang et al. [52] reported that clinicians turned to colleagues to receive what they called social decision support. In another publication, Yang et al. also reported that seasoned physicians indicated that their *"dream DST should play a role similar to a mid-level clinician"* (p. 238) and that it should provide additional information for decision-making [53]. A need for relatedness – experiencing a sense of belonging, attachment, and closeness – could underlie such statements. In anesthesia, where many decisions and actions are performed in teams (including anesthetic nurses, anesthesiologists, supervising anesthesiologists, and often even members of other clinical specialties), the psychological need for relatedness may be of great importance. In summary, a needs-based UX perspective may explain described issues in prior work and could, therefore, offer the potential for future DST designs that fit better into the context.

An embodied perspective on UX emphasizes the user's intentionality, the physical coupling of the user and the environment, and the context of the interaction [10]. The embodied perspective therefore especially highlights the user's physical body as being situated in a specific physical environment and social context [20]. Concerning UX, embodiment suggests that a "good" UX is a fluent interaction with the social and physical environment. Van Dijk and Hummels [48] suggested considering a skillful coupling of the user with the environment and social coordination when designing interactive technology. In the literature, we identified different examples of environmental coupling. Less fluent interaction was reported when interaction times were extended due to the required data entries [39]. A positive example of environmental coupling can be found in the work of Yang et al. [53], who carefully integrated DST information in the workflow of clinicians and therefore achieved a fluent interaction with the environment (and potentially a better UX).

The present study

In the following study, we aimed to test the benefits of a UX theory-driven design method by designing a tool to support anesthesiologists while making decisions on diagnoses and treatments during operating room crises. To this end, we followed a contextual design process [24] with an emphasis on needs-based and embodied UX in the context of operating room crises in anesthesia. The contextual inquiry resulted in a set of key insights and visions that further guided the iterative design of the DST that we named "Cassandra". The name is a reference to Greek mythology. If people had believed in Cassandra's ability to predict the future, considerable damage could have been prevented. We conducted a pilot evaluation of Cassandra using a full-scale, high-fidelity medical simulation, including anesthetic teams. Subsequently, we conducted semi-structured interviews with participants. We analyzed the interviews and videos of each scenario to evaluate whether the DST Cassandra fulfilled the key insights generated from the contextual inquiry. We discuss the preliminary results of the study in the context of the suggested UX design approach to DSTs and the general DST literature.

DESIGN OF CASSANDRA

Anesthesia in the operating room context

In many operations, the anesthesiologist's task is to induce and maintain general anesthesia into a patient. This includes rapidly, safely, and pleasantly producing and maintaining amnesia, analgesia, akinesia, and an autonomic and sensory block while maintaining hemodynamic stability and sufficient ventilation [28]. In general, the anesthetic team works in a non-sterile environment, which is separated from the sterile surgical side by a drape (see Figure 1). In the hospital under study, the anesthesiologist is supported by an anesthetic nurse and may call for an additional supervising senior anesthesiologist.

Contextual Design

To understand the anesthesiologists' work in context, we followed a contextual design approach [24] with a particular emphasis on psychological needs and environmental and social coupling. Because intraoperative crises are unpredictable and fortunately infrequent, a usual contextual inquiry was not possible. Instead, we conducted four semi-structured interviews with three senior anesthesiologists (work experience >5 years each) and one junior anesthesiologist (work experience <5 years). Additionally, we inspected three videos of simulator-based crisis training with anesthesiologists with one year of work experience. Finally, we observed and participated as actors in three simulator-based crises-training scenarios with twelve medical students in their anesthesia traineeship.

All participants were recruited from the University Hospital of Würzburg. Four researchers consolidated interview and observation data into affinity notes. The notes were then structured in an affinity diagram and grouped into seven categories. All researchers performed a wall walk along the affinity diagram in order to generate design ideas and key insights from the data [cf. 24]. The key insights served as requirements for the further design process.

We identified three groups of key insights that also reflected our theoretical orientation of a needs-based and embodied UX. The first group, the *need for autonomy and competence,* included four insights: (1)The system should not patronize its users, (2) the system should respect different levels of experience, (3) the system should be trusted and accepted by users, and (4) the system should not make users feel guilty in any way. In other words, the system should not make users feel that they should have already known displayed information and that the system displayed the information just because the users' performance was not 'good' enough.

The second group, the *environmental context*, contained three insights: (1) The system should support clinicians in both calm and stressful situations, (2) the system should only require minimal interaction to avoid additional stress for clinicians, and (3) the system should be able to handle different kinds of data (e.g., continuous vital sign parameters or specific patient information).

The third group, *teamwork and communication*, contained two insights: (1) The system should enable all team members to have the same information about the patient and should, therefore, be accessible for the whole team and (2) the system should support the joint process of decision-making.

Based on these insights, visions were sketched. All 22 visions were clustered in three topics: *teamwork*, *display and interaction concept*, and *finding a diagnosis*. We briefly report the most influential vision (i.e., satisfying the most key insights) of each category. To improve teamwork, we envisaged placing the system in front of the drape separating anesthesiologists from surgeons, which would make the system accessible for all anesthetic team members at all times. Concerning the display concept, we envisaged the system as a mirror reflecting the patient's body, allowing vital signs to be presented close to the patient's body. In relation to finding a diagnosis, we considered an advocatus diaboli, who would always aim to falsify a diagnosis that was entered by an anesthesiologist. All visions were evaluated against the key insights. The final prototype, Cassandra, is based on a combination of different visions and we considered as many key insights as possible.

Iterative Development

We developed a first concept and ran three iterations of user tests and interviews. In each iteration, an interviewer showed the prototype to one user at a time, and another researcher took notes. Users were asked to think of an imaginary scenario of a standard anesthesia procedure or intraoperative crisis they recently experienced during work and browse through the prototype commenting on the interaction and what should be improved. Overall, one human factors specialist, six senior anesthesiologists (work experience >5 years each), two junior anesthesiologists (work experience <5 years each), and one anesthetic nurse tested different versions of the prototype. The first version was a paper prototype, and versions two and three were digital wireframe prototypes presented on a laptop (see supplemental material Figure 1-3 for examples). After each session with users, four researchers consolidated the issues and emerging requirements, which were implemented before the next iteration.

At the same time, three anesthesiologists developed and iterated on a set of symptoms and their relations to 15 diagnoses to establish the content for a test scenario. All diagnoses shared symptoms with malignant hyperthermia, which is a rare, potentially life-threatening incident with non-specific early symptoms that could easily be confused with many differential diagnoses [42].

Final Prototype of Cassandra

While Cassandra is not activated, the center of the screen shows a video stream of the surgical side. This "window" should support situation awareness, communication, and teamwork between the anesthetic and surgical teams. On the right side, the screen always displays basic patient information and their vital signs (see Figure 2). The anesthesiologist does not need to turn away from the patient, as is necessary in current standard operating rooms, which should enable environmental coupling. To satisfy the anesthesiologists' need for autonomy, we decided to only display the diagnosis tool on demand. To access the core functionality of Cassandra, only one touch on the display is required, and the tentative diagnosis may be entered using a Bluetooth keyboard or speech recognition to further enable environmental coupling.

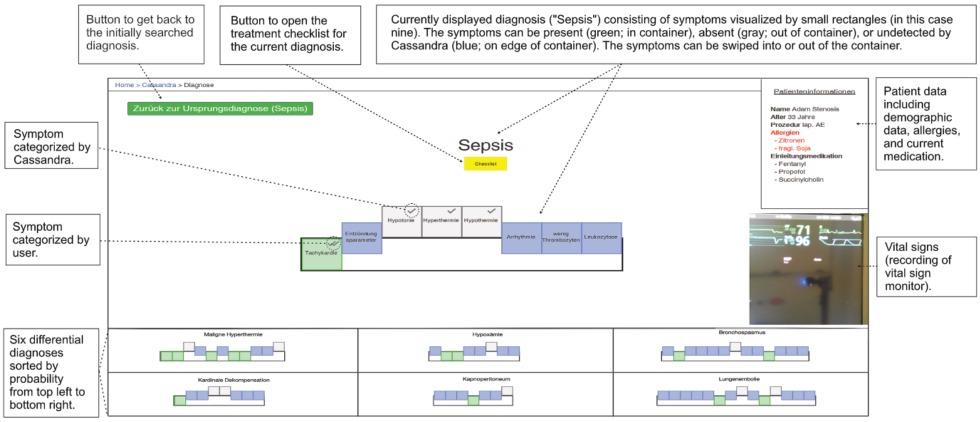

Figure 2. Screenshot of the final Cassandra prototype with explanations. Cassandra shows the diagnosis "Sepsis" and the most likely differential diagnoses for a simulated scenario.

Figure 2 shows the core functionality of Cassandra. The entered tentative diagnosis is displayed in the center of the screen. In the lower part of the screen, the six most likely differential diagnoses are displayed in order of probability. This enables users to consider alternatives, and they may therefore be less likely to be affected by premature closure (accepting one diagnosis instead of considering reasonable alternatives) or confirmation bias (the tendency to seek supporting evidence rather than trying to falsify the tentative diagnosis). Independent of experience and of whether an entered diagnosis was 'correct', users are enabled to identify the most probable diagnosis. However, the user remains in charge of deciding which diagnosis to consider, satisfying their need for autonomy and competence.

Diagnoses are visualized by a "container bar" consisting of boxes that represent all symptoms of a diagnosis. These can be present (green; in the container), absent (grey; out of the container) or undetectable by Cassandra (blue; on edge of the container). Because the symptoms are arranged by importance concerning the respective diagnosis from left to right, Cassandra can be used as a cognitive aid for which symptoms to check and in which order (reducing cognitive biases of omission and strengthening competence). Cassandra automatically updates the symptoms' states and differential diagnoses based on all digitally available data (satisfying the requirement for minimal interaction and, therefore, tight environmental coupling). Present and absent symptoms include checkmarks that indicate that a symptom was judged by Cassandra (single checkmark) or by a user (double checkmark), supporting the need for autonomy. Anesthesiologists may change the state of each symptom by swiping the symptoms in or out of the diagnosis container according to their assessment. This functionality fulfills the need for autonomy and competence.

The diagnosis visualization acts as a shared knowledge base for the team enabling rapid evaluation of the tentative diagnosis. The differential diagnoses are displayed as miniature versions to help the team to consider alternatives and to find the right diagnosis even if a wrong diagnosis has initially been entered. For example, Figure 2 shows that the first differential diagnosis ("Malignant Hyperthermia") may be more likely due to more present symptoms (i.e., green boxes) than the entered primary diagnosis ("Sepsis"). Finally, the visualization of the diagnosis in Cassandra provides a basis for explaining the situation to staff who have been called for help. If the anesthesiologist decides on a diagnosis, the respective treatment information can be retrieved. Treatment information is based on existing checklists (a translated version of the Stanford manual [14]) presented as a list that can be checked off to help allocate human resources.

To realize the final prototype of Cassandra, we set up a mySQL database for storing diagnoses, symptoms, and respective treatments. The connection to the database was established via Node.js. The main application was implemented with HTML, CSS, and JavaScript and optimized for use on large touchscreen displays.

PILOT EVALUATION OF CASSANDRA

We preliminarily evaluated Cassandra within a simulator-based crisis scenario on malignant hyperthermia (see Figure 1). A high-fidelity, full-scale patient simulator closely resembles realistic circumstances of an ongoing procedure in the operating room and is, therefore, an appropriate method to evaluate new technology [9]. The aim of the pilot evaluation was to investigate whether the design of Cassandra could fulfill the key insights gathered during the contextual design.

Participants

In total, nine participants (four anesthetic nurses, five anesthesiologists) were recruited from the University Hospital of Würzburg. Participants took part in teams consisting of one nurse and one anesthesiologist (and in one case two anesthesiologists). The average age of the participants was 31 years (SD = 6.53) and their working experience varied between 0.5 and 13 years (M = 6.2, SD = 4.8). Participants received a five Euro cafeteria voucher. All participants had taken part in simulation-based training previously. None of the participants had experienced a case of malignant hyperthermia in real life. None of the participants had seen previous versions of Cassandra.

Procedure and Materials

Participants arrived, were informed, gave consent, and received a short, standardized introduction to familiarize them with Cassandra. Next, the participants left the simulation room, and the simulation started. Each scenario included two confederates acting as surgeons who simulated a laparoscopic appendectomy in a high-fidelity patient simulator (HPS Human Patient Simulator; CAE Healthcare, Sarasota, FL, USA) and one confederate anesthesiologist. The participants were introduced into the scenario via a handover from the confederate anesthesiologist who then left the simulation room and did not re-enter the scenario. After the handover, the previously stable patient soon started to deteriorate, and participants were required to provide diagnosis and treatment.

After the scenario, we conducted a semi-structured interview that aimed to understand the participants' experience in detail, asking questions like *"How did the system fit into your workflow and your thought processes during the operating room crisis?"*. Although we prepared an initial set of questions (see supplemental material Table 1), we were interested in any aspect that the participants raised during the interview. All interviews were audio-recorded, and the overall session took around 45 to 60 minutes.

Cassandra was presented on an Iiyama T2735 touchscreen (13.2" x 23.5"), which was positioned in front of the green drape (see Figure 1). During each scenario, two anesthesiologists controlled the patient simulator from an adjoining control room. A researcher in the control room controlled Cassandra in a wizard-of-oz manner. The researcher imitated the automatic detection of changing symptoms according to the state of the patient simulator. To realize the "window" functionality and to provide the current vital sign parameters, the patient monitor and a section of the surgical side were recorded with two webcams (Logitech HD Pro Webcam C920) and displayed in real-time at the intended positions within Cassandra. All scenarios were audio-video recorded using two cameras (GoPro HERO4®).

Results and Discussion of the Pilot Evaluation

Video data

The videos of each simulation session were reviewed independently by two researchers, and sequence models of all interactions with the system were developed and consolidated across all participants [24]. This procedure resulted in four general interaction models with Cassandra (see Figure 3 and supplemental materials Figures 4-6): (1) context with Cassandra, (2) calling Cassandra for help, (3) finding a diagnosis, and (4) checklist use.

First, with Cassandra in the context, anesthesiologists relied on tools they were already familiar with (monitors, ventilator, and paper-based patient information) during patient hand-over. Most participants first interacted with Cassandra before the patient deteriorated by observing the surgery side through Cassandra's "window", especially if surgeons had questions or caused unclear sounds. Participants also frequently looked back and forth between the vital sign monitors and Cassandra.

Second, concerning calling Cassandra for help, most teams needed a hint (delivered by an instructed surgeon) to start interacting with Cassandra during the operating room crisis. Both speech and keyboard input was used to enter a diagnosis. The aspect of not initiating the interaction even if the system detects anomalies and instead waiting for user input of a tentative diagnosis was intended to preserve the user's autonomy. However, the question arises regarding whether this emphasis on autonomy (in combination with users possessing little experience and an operating room crisis) might have lead to non-use and should, therefore, be avoided, or whether this handling is vital for the general acceptance of Cassandra and should be preserved.

Another example of need satisfaction may be found in the analysis of the inserted tentative diagnoses. Experienced anesthesiologists entered the most probable diagnosis as the first tentative diagnosis, while less experienced anesthesiologists did not. Nonetheless in both cases, the anesthesiologist was able to identify the most probable diagnosis by being confronted with several differential diagnoses compared to the diagnosis entered (see Figure 2, in which the first differential diagnosis is most probable).

While entering a tentative diagnosis allows anesthesiologists to stick with their standard way of thinking during an operating room crisis (type 1 thinking: quick, intuitive, error-prone [38]), this input activates Cassandra, which displays structured information on differential diagnoses (type 2 "thinking": slow, analytical [38]). Therefore, both ways of thinking become seamlessly interwoven through the interaction between anesthesiologists and Cassandra, which subsequently may prevent typical cognitive biases such as premature closure (accepting one diagnosis instead of considering reasonable alternatives) or omission bias (the tendency towards inaction rather than action). The above aspect eventually supports the anesthesiologists' competence by strengthening them against common cognitive biases in anesthesia [45] and thereby enhancing their performance.

Third, regarding finding a diagnosis, anesthesiologists used various interaction patterns (see Figure 3). Again

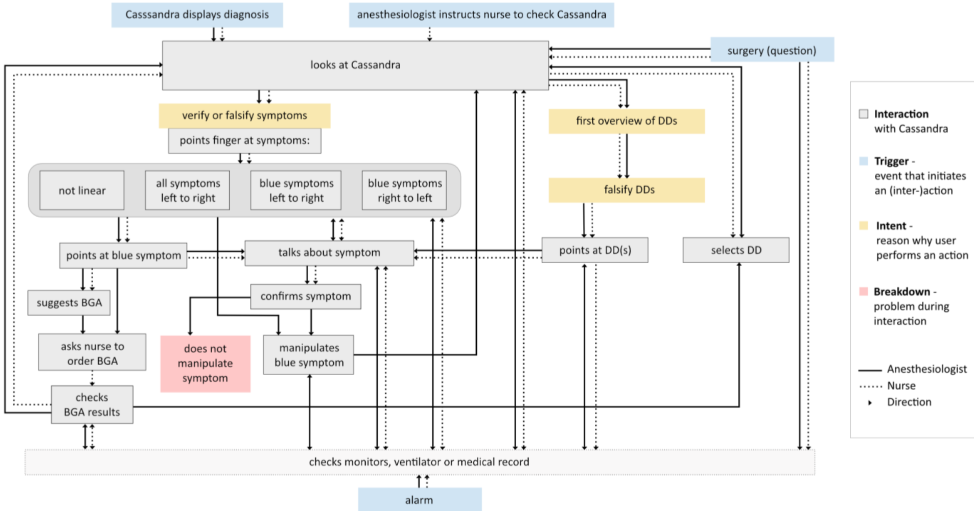

Figure 3. Consolidated sequence model of the interaction model 'finding a diagnosis'. The consolidation reveals the flexibility when using Cassandra. Explanations of the different pathways are provided in the text.

demonstrating that Cassandra supported their need for autonomy, anesthesiologists used Cassandra to obtain an overview of potential diagnosis and to validate the symptoms by pointing at the displayed symptoms, reading the symptoms aloud, and confirming the symptoms using the monitor and ventilator.

For the actual diagnostic process, participants used different strategies to validate the symptoms of a single diagnosis. Most participants linearly reviewed the symptoms, either starting from the left or only inspecting the blue symptoms (i.e., undetectable by Cassandra) from left to right or right to left. Additionally, Cassandra was consulted after an alarm or when a vital sign parameter changed. No participant used the option to mark system-checked symptoms with another checkmark to indicate a verification by the user. Prompted by blue symptoms, which could only be assessed with the results of a blood gas analysis, all teams requested a blood gas analysis and subsequently reevaluated the diagnosis by reiterating through the symptoms. This example also demonstrates another way in which Cassandra supported autonomy and competence. Cassandra does not make any decision but only provides information. Therefore, Cassandra allows the users to decide on which actions to perform and thus leaves the perceived locus of competence with the anesthetic team. Anecdotally, the two anesthesiologists controlling the simulation in each scenario were surprised by the early demand for a blood gas analysis.

During the reevaluation of the diagnosis upon reception of the blood gas analysis, teams distributed their tasks so that, for example, the nurse would examine the symptoms' state in Cassandra while the anesthesiologist would read the results of the blood gas analysis out loud. By allowing equal access, we considered Cassandra to not promote hierarchies in the clinical context. Nonetheless, existing hierarchies led to situations where anesthetic nurses never operated Cassandra or did only so upon the instructions of an anesthesiologist. Such an observation may be interpreted in two ways. First, the design did not dictate a single way of usage but also allowed Cassandra to be integrated into existing (hierarchical) workflows. Second, while some anesthetic nurses felt empowered to speak out, others stuck with known procedures. Change was, therefore, not enforced but welcomed by design, which allowed autonomy and individual decision-making on how to use the system. This may be seen as a compromise between enforcing change by design and embedding a system into existing contexts.

In all scenarios, only blue symptoms were manipulated, and in one case, no symptom was manipulated at all. Moreover, Cassandra was used to obtain an overview of differential diagnoses and to consider and select more suitable differential diagnoses. Although the differential diagnoses were sorted by probability, some anesthesiologists overruled the system based on their knowledge and competence by not selecting, according to Cassandra, the most likely diagnosis.

Fourth, the checklist was only used from the second scenario onwards after a more salient button was added. Nevertheless, users hesitated in the interaction with Cassandra when switching back and forth between the checklist and the diagnosis. Analogous to the interaction with Cassandra while deciding upon a diagnosis, the checklist was read aloud, pointed at, and checked against the monitors and ventilator. Subsequently, anesthesiologists ticked off the completed checklist items.

During all phases of the simulated surgery, we were able to observe many interactions between users and Cassandra, including watching, pointing, and discussing information displayed within the team, even though the minority of interactions entailed physical manipulation with the system.

Interview data

In order to analyze the interviews, we thematically coded single interview quotes following the content-structuring qualitative analysis as described by Mayring [32]. The procedure resulted in five themes based on the codes: (1) the need for autonomy, (2) the need for competence, (3) the need for relatedness and social coupling, (4) insights concerned with the integration into the environment, and (5) usability-related issues. We supplement our findings with quotes referencing participants' ID (number) and profession (A = anesthesiologist and N = anesthetic nurse).

First, concerning autonomy, one participant stated that in his opinion, *"the decision-making process can´t be taken away from the anesthesiologists. The system narrows down possibilities, and maybe you discover something that you didn´t consider before, but the decision in which direction you want to go is not taken away by the system. I wouldn´t like it if it did [take the decision away]"* (A3). The statement indicates that our system supported rather than patronized clinicians during usage. The system kept the level of self-involvement of the participants high.

One clinician, however, criticized that Cassandra lacked the functionality to verify the origin of the data used. He was concerned with the symptoms the system detected: *"[...] I wouldn´t know where the information is coming from. Is the [laboratory value] from [...] five years ago or [recent] ...? We have many systems displaying incorrect information to us. Cassandra should, in some way, be able to display where the data is coming from [...]. [However,] I wouldn´t like [Cassandra] to display this [information] permanently."* (A8). The above-described fear that the system may contain false information was reinforced by the electronic anesthesia patient record currently in use, as it may contain outdated information: *"Especially with laboratory data, you have, for example, [values] from 2002."* (N9). The missing information about the data's origin may be interpreted as reducing the quality of self- involvement and suggests ways to improve autonomy. For example, data sources for each piece of information could be made visible on demand. As Cassandra did not yet support such an action, participants felt compelled to manually compare the values displayed on Cassandra and the values displayed on standard devices (as observed in the video recordings).

Second, regarding competence, four of the participants responded that Cassandra made them feel like they had considered everything essential. One said: *"I saw hyperkalemia in one of the boxes. You need a blood gas analysis for that. [Cassandra] is, of course, a mnemonic for these actions."* (A1). Another participant said: *"It´s like a checklist for pilots [with things] that you should know, but still you go through so that you don´t forget something. That you didn´t miss something in the hectic situation."* (A3). Two participants said that Cassandra helped them by making them *"[...] see [their] thoughts visualized"* (A1). The quotes complement the findings from video analysis and indicate that the system supported the participants to perform well and thereby to perceive competence.

When asked whether Cassandra should also contain anesthesia management problems such as a leaking respiratory mask, two of the participants answered yes, one of them said: *"If you could enter 'show me respiratory problems' [...] you would get the possibility, as a beginner, to [...] work on a list of things as an expert would do [...] [You] get a list of things that helps you [by telling you what to check now]."* (A3). Two more experienced anesthesiologists did not see value in adding these problems, saying: *"When it is a small complication [...], I don´t need [help]"* (A8).

When asked about the differential diagnoses display, three participants described it as interesting and helpful but stated that the display did not influence them during the decision-making processes, which is not supported by video analysis (e.g., the early demand for blood gas analysis). Another three participants stated that they felt strongly influenced by the differential diagnoses display. Three further participants had either positive, mixed, or negative feelings concerning the display. The participant with mixed feelings explained: *"When you read one of them [one of the diagnoses], you might then have it in your head and can´t let go of it while it might actually be a different [diagnosis]. This might lead to a restricted angle of view. On the other hand, it can, of course, be a thought-provoking impulse"* (N5). Thus, some participants were afraid that the differential diagnoses display might interfere with their decision-making process, harming their ability to identify the correct diagnosis while others valued the display as beneficial. In other words, some participants stated the risk of fixating on one diagnosis. Video analysis did not reveal such issues (to the contrary, anesthesiologists switched back and forth between differential diagnoses), and Cassandra was designed to minimize cognitive biases.

The results of our first pilot exploration were encouraging, given that all teams were able to determine the correct diagnosis. Even so, a future study needs to explore whether unwanted fixations take place when using Cassandra with its current visual appearance of the diagnosis display.

Third, considering the need for relatedness and social coupling, Cassandra was of great value to the participants. Six participants said that Cassandra improved their communication or their team integration within the anesthetic team, one participant said that the communication *"is better, for sure. [Cassandra helped] to find a joint we-know"* (A3), while another participant stated that he thinks, *"that you speak out loud your thoughts more frequently"* (A6). Another participant said: *"you can use it as a communication tool that everyone knows what is going on […] I am currently missing such a possibility, to be honest"* (N7). We interpret this quote as a satisfaction of the need for relatedness. Participants experienced a sense of belonging and closeness that was induced by increased communication behavior. The above quotes also support the idea that the system increased social coupling – fluent interaction between the users. Whether Cassandra led to increased communication among the members of the anesthetic team is an interesting aspect that should be explored in a comparative follow-up experiment.

Cassandra failed, however, to improve communication between the anesthetic and surgical teams, as our concept (the "window" to the operating side) was seen as dispensable by eight participants. However, video analysis revealed that the window was nevertheless used, especially if surgeons had questions or caused unclear sounds. When asked whether it would help the surgical team to have a video feed of the anesthetic side provided, four participants responded that this would be unnecessary as well. No participant valued a "window" in both directions. The participants stated that the step stool that is currently used to look over the drape separating the surgery and the anesthetic team enhances communication more than the video solution.

Fourth, regarding the environmental coupling, participants' statements varied. One participant said that the positioning of the system was good, compared to five participants who stated that it was not optimal, mainly for practicality reasons. One such reason, for example, was the (interpreted) need to set it up before and remove it after each surgery. Because set up and removal would require additional work, anesthesiologists and especially anesthetic nurses indicated that the system would not be used. The main problem suggested by the anesthesiologists, however, was that the surgeons would probably complain about Cassandra being in this position, as it might interfere with some surgical procedures or emergency interventions. When not considering setup and interference with interventions, three participants rated the position as positive. Two participants advised that the screen be mounted on a suspension while two participants suggested it be placed next to the patient, for example, closer to the monitoring, because *"out of habit, you always look at the monitoring anyway"* (N7). While the positioning of Cassandra led to several positive aspects such as equal access and reported increased communication, participants were concerned about the position for practical reasons. As can be seen in Figure 1, the Cassandra prototype was rather bulky, being mounted on a wooden frame and including a massive touch display. Other technologies such as a flexible, see-through display, or a projection of Cassandra on the drape might be more suitable and might mitigate the stated concerns about additional work and interference with interventions. The videos and parts of the interviews revealed another aspect of participants' skepticism about the position of Cassandra. By duplicating and displaying the current vital signs of the patient within Cassandra, we aimed to bring vital sign information closer to the patient so that turning away from the patient would become unnecessary for anesthesiologists. However, out of habit, all of our participants used the usual vital sign monitoring to check vital signs. It would be interesting to see how anesthesiologists would integrate the vital sign view of Cassandra into their workflow when establishing new routines over time.

Fifth, concerning usability, participants valued the general usability of the system. Five participants noted that the system was intuitive to use. Five participants said that the system was structured and not cluttered. However, participants rarely used the checklist to treat a patient, which could be related to a design issue. During the first scenario, participants had to click on the diagnosis's name to get to the treatment screen. Both participants stated that they forgot about the existence of this function. For the second scenario, a bright yellow button led to treatment information. Still, three participants said that the treatment checklist was hard to find, as the button was rather small and placed not on the bottom right side but the center, right below the name of the diagnosis. Overall, Cassandra was usable during calm and stressful situations, and the interaction was perceived as being natural and fluent (except for finding the checklists for treatment, which needs further improvement).

GENERAL DISCUSSION

Overall, the results of our pilot evaluation suggested that need fulfilment in clinical contexts can be vital for DST acceptance. Participants appreciated Cassandra when it was able to support need fulfilment (e.g., satisfied need for relatedness through an increase of team communication triggered by the system) but rejected Cassandra if it did not (e.g., lack of satisfaction for autonomy when data sources were unverifiable in the system). Concerning the social and environmental coupling, Cassandra allowed the team to fluently work towards a joint goal by requiring only minimal interaction, by allowing to efficiently distribute tasks, and by its affordance to gather around the display. While Cassandra and its placement created a setting that served environmental and social coupling, participants in the pilot evaluation relied on their habits as well (e.g., turned away to check paper lists) which interrupted the new setting. Environmental and social coupling can only vaguely be judged based on this short-term pilot evaluation; however, our results seem promising.

UX has so far seldom been considered in the context of safety-critical domains [e.g., 33, 41] and has even been considered to not be relevant [18, 34]. Whether designed for UX or not, DSTs provide clinicians with (user) experiences through interaction with them, that may be perceived as good or bad [20] depending on the degree of psychological need fulfillment [21] or on the fluency of environment integration

[48]. This general discussion aims to situate our and previous findings within UX theory.

We identified that the reported failures of DSTs might potentially be related to failures of the psychological need satisfaction of clinicians when interacting with DSTs, such as autonomy [47, 51] and competence [6, 47, 49, 52]. Our contextual inquiries complemented these findings by indicating that psychological needs are essential for DSTs in the context of anesthesia, and we therefore placed a strong emphasis on need satisfaction in the design. For instance, a design that emphasizes autonomy and competence may leave any decision to its users by providing relevant information and allowing users to make well-informed decisions on their own. In doing so, the design may promote the skills of the users rather than de-skilling them by introducing DSTs [11] and in the clinical context, it may make *"clinicians feel they are doing better work, and not necessarily automating the part of the work that makes them feel like an expert"* [53]. We think that this will become even more important with recent research on the so-called precision medicine of DSTs that aims to personalize clinical care based on big data [26]. Due to complexity, the expertise is "allocated" to the DSTs, and careful interface design might be necessary to still support psychological need fulfillment and the subsequent acceptance of DSTs.

The importance of need fulfillment became even more apparent in Cassandra's lack of a function to determine data sources. First, clinicians are experts and (like any expert) want to experience their expertise. Not being able to trace back and double-check information and understand the reasoning of a DST does not support psychological need fulfillment and does not motivate clinicians to use a DST [53]. Second, as our results indicated, not being able to determine the data source of information can even lead to distrust. Alexander and Joshi [1] stated that one of the anesthesiologists' most significant challenges would be validating the safety and efficacy of such systems. Similar to the comment that DSTs should be like mid-level clinicians and support decision-making with additional information [53], Alexander and Joshi stated that DSTs need to be observed like junior residents to ensure safety and efficacy [1]. Questioning and re-evaluating data is, therefore, an essential task of anesthesiologists, and DSTs should provide the information and means to accomplish this task. In summary, especially when professionals are required to make use of technology, such usage will more likely be successful if it provides them with additional, valuable skills that help them to overcome human cognitive limitations and therefore enhances users' competence. This is particularly true in the present example because, in the end, clinicians are the ones in charge, and technology should support them in improving their performance.

From an embodied UX perspective [48], technology can be successful if it can be integrated into existing social and contextual structures and enable a fluent interaction. In the case of anesthetic operating room crises, for example, anesthesiologists need to receive information at the right time and place, which is prepared and presented in a way that is suitable to their workflow and considers teamwork within anesthesia. Trying to adapt Cassandra to the context of teamwork in anesthesia, our design aimed to foster a more structured communication within the team. Consistent with previous research that used large displays [16, 51], Cassandra seemed to serve as a tool to consult, to evaluate one's thoughts, and as a trigger to share one's thoughts. Further aspects related to the embodied UX perspective are time and resource restrictions that frame the interaction with a DST. While entering data into a DST and long interaction times [39] would not support fluent interaction and environmental coupling during operating room crises, Cassandra required only minimal interaction and enabled joint, flexible use, which was accepted by our participants. Designing for minimal interaction times and eased access seems essential for a DST's fit into (simulated) operating room crises.

Limitations and Future Work

The present study has several limitations. First, although small sample sizes are standard in contextual design [24], our sample size in both qualitative explorations and especially the pilot evaluation was rather small. Second, the pilot evaluation focused on qualitative feedback, and we did not evaluate a UX-driven design vs. a non-UX-driven design. Future research should use such an experimental comparison and use quantitative measures concerning psychological need satisfaction [e.g., 21]. Third, the pilot evaluation was conducted within a simulated medical environment. Fourth, we only considered a set of 15 diagnoses related to the scenario in the pilot evaluation, and future research is needed to investigate the scalability of Cassandra to more than one scenario. Fifth, besides UX, there are certainly other approaches to increase the environmental fit and acceptance of DSTs. Previous work on automation can provide valuable input here [30] as, for example, automation exposure has been found to increase the acceptance of automated processes [3].

CONCLUSION

By taking a needs-based and embodied UX perspective when designing a DST, our approach differed from most related DST literature, which focused on interaction concepts related to safety and efficiency. This new perspective allowed us to (1) explain previously identified DST implementation and acceptance problems and *why* some design solutions were more successful and (2) generate new insights in the design and use of DSTs in anesthesia. We believe that the UX perspective may also contribute to DST design beyond the clinical context and should be considered in future DST design.

ACKNOWLEDGEMENTS

We thank our clinical participants whose commitment and enthusiasm made the present work possible.

REFERENCES

[1] John C. Alexander and Girish P. Joshi. 2018. Anesthesiology, automation, and artificial intelligence. In *Baylor University Medical Center Proceedings*. Taylor & Francis, 117-119. https://doi.org/10.1080/08998280.2017.1391036

[2] Edouard Battegay and Stefano Bassetti. 2017. *Grundlagen der Differenzialdiagnose.* In: Differenzialdiagnose Innere Krankheiten, Georg Thieme Verlag, New York, NY, USA, https://doi.org/10.1055/b-004-129980

[3] Marek Bekier and Brett R. Molesworth. 2017. Altering user' acceptance of automation through prior automation exposure. *Ergonomics*, 60, 6: 745-753. https://doi.org/10.1080/00140139.2016.1216610

[4] Casey C. Bennett and Kris Hauser. 2013. Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial intelligence in medicine*, 57, 1: 9-19. https://doi.org/10.1016/j.artmed.2012.12.003

[5] Eta S. Berner and Tonya J. La Lande. 2007. *Overview of Clinical Decision Support Systems.* In: Clinical Decision Support Systems: Theory and Practice, Springer, Cham, Germany, 3-22. https://doi.org/10.1007/978-0-387-38319-4_1

[6] Maxime Cannesson and Joseph Rinehart. 2014. Closed-loop systems and automation in the era of patients safety and perioperative medicine. *Journal of clinical monitoring and computing*, 28, 1-3: https://doi.org/10.1007/s10877-013-9537-3

[7] Basit Chaudhry. 2008. Computerized clinical decision support: will it transform healthcare? *Journal of general internal medicine*, 23: 85-87. https://doi.org/10.1007/s11606-007-0432-9

[8] Keng Sheng Chew, Jeroen van Merrienboer, and Steven J. Durning. 2017. Investing in the use of a checklist during differential diagnoses consideration: what's the trade-off? *BMC medical education*, 17, 1: 234. https://doi.org/10.1186/s12909-017-1078-x

[9] David Cumin, Jennifer M. Weller, Kaylene Henderson, and Alan. F. Merry. 2010. Standards for simulation in anaesthesia: creating confidence in the tools. *British journal of anaesthesia*, 105, 1: 45-51. https://doi.org/10.1093/bja/aeq095

[10] Paul Dourish. 2004. *Where the action is: the foundations of embodied interaction*. MIT press, Cambridge, MA, USA.

[11] Pelle Ehn. 1993. *Scandinavian design: On participation and skill.* In: Participatory design: Principles and practices, 41-77.

[12] Glyn Elwyn, Isabelle Scholl, Caroline Tietbohl, Mala Mann, Adrian Edwards, Catharine Clay, France Légaré, Trudy van der Weijden, Carmen L. Lewis, Richard M. Wexler, and Dominick L. Frosch. 2013. "Many miles to go…": a systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC medical informatics and decision making*, 13: https://doi.org/10.1186/1472-6947-13-S2-S14

[13] Aude Fourcade, Jean-Louis Blache, Catherine Grenier, Jean-Louis Bourgain, and Etienne Minvielle. 2012. Barriers to staff adoption of a surgical safety checklist. *BMJ Quality & Safety*, 21, 3: 191-197. http://dx.doi.org/10.1136/bmjqs-2011-000094

[14] David M. Gaba, Kevin J. Fish, and Steven K. Howard. 1994. *Crisis management in anesthesiology*. Churchill Livingstone, New York, NY, USA.

[15] Atul Gawande. 2009. *The checklist manifesto: How to get things right*. Metropolitan Books, New York.

[16] Michael J. Gonzales, Joshua M. Henry, Aaron W. Calhoun, and Laurel D. Riek. 2016. Visual task: a collaborative cognitive aid for acute care resuscitation. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 45-52. http://dl.acm.org/citation.cfm?id=3021319.3021327

[17] Mark L. Graber, Nancy Franklin, and Ruthanna Gordon. 2005. Diagnostic error in internal medicine. *Archives of internal medicine*, 165, 13: 1493-1499. https://doi.org/10.1001/archinte.165.13.1493

[18] Jonathan Grudin. 2017. From tool to partner: The evolution of human-computer interaction. *Synthesis Lectures on Human-Centered informatics*, 10, 1: https://doi.org/10.2200/S00745ED1V01Y201612HCI035

[19] Tobias Grundgeiger, Stephan Huber, Daniel Reinhardt, Andreas Steinisch, Oliver Happel, and Thomas Wurmb. 2019. Cognitive Aids in Acute Care: Investigating How Cognitive Aids Affect and Support In-hospital Emergency Teams. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1-14. http://doi.acm.org/10.1145/3290605.3300884

[20] Tobias Grundgeiger, Joern Hurtienne, and Oliver Happel. 2020. Why and How to Approach User Experience in Safety-Critical Domains: The Example of Healthcare. *Human Factors*: https://doi.org/10.1177/0018720819887575

[21] Marc Hassenzahl, Sarah Diefenbach, and Anja Göritz. 2010. Needs, affect, and interactive products – Facets of user experience. *Interacting with computers*, 22, 5: 353-362. https://doi.org/10.1016/j.intcom.2010.04.002

[22] Thomas M. Hemmerling, Fabrizio Cirillo, and Shantale Cyr. 2012. *Decision support systems in medicine-anesthesia, critical care and intensive care medicine.* In: Decision Support Systems, IntechOpen, London, UK, http://dx.doi.org/10.5772/51756

[23] David L. Hepner, Alexander F. Arriaga, Jeffrey B. Cooper, Sara N. Goldhaber-Fiebert, David M. Gaba, William R. Berry, Daniel J. Boorman, and Angela M. Bader. 2017. Operating room crisis checklists and emergency manuals. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 127, 2: 384-392. https://doi.org/10.1097/ALN.0000000000001731

[24] Karen Holtzblatt and Hugh Beyer. 2017. *Contextual Design: Design for Life*. Morgan Kaufmann, Burlington, MA, USA.

[25] Monique W. Jaspers, Marian Smeulers, Hester Vermeulen, and Linda W. Peute. 2011. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *Journal of the American Medical Informatics Association*, 18, 3: 327-334. https://doi.org/10.1136/amiajnl-2011-000094

[26] Robert A. Jenders. 2017. Advances in clinical decision support: Highlights of practice and the literature 2015-2016. *Yearbook of medical informatics*, 26, 1: 125-132. http://doi.org/10.15265/IY-2017-012

[27] Kensaku Kawamoto, Caitlin A. Houlihan, Andrew Balas, and David F. Lobach. 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*, 330: 765. https://doi.org/10.1136/bmj.38398.500764.8F

[28] A. King and L.M. Weaving. 2017. General anesthesia: Induction. Retrieved September 10, 2019 from https://www.uptodate.com/contents/general-anesthesia-induction?source=search_result&search=general%20anesthesia&selectedTitle=3~150.

[29] Leah Kulp, Aleksandra Sarcevic, Megan Cheng, Yinan Zheng, and Randall S Burd. 2019. Comparing the Effects of Paper and Digital Checklists on Team Performance in Time-Critical Work. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1-13. http://doi.acm.org/10.1145/3290605.3300777

[30] John D. Lee, Christopher D. Wickens, Yili Liu, and Linda Ng Boyle. 2017. *Human Automation Interaction.* In: Designing for people: An introduction to human factors engineering, CreateSpace, Scotts Valley, CA, USA, 357-387.

[31] Stuart Marshall. 2013. The Use of Cognitive Aids During Emergencies in Anesthesia: A Review of the Literature. *Anesthesia Analgesia*, 117, 5: 1162-1171. https://doi.org/10.1213/ANE.0b013e31829c397b

[32] Philipp Mayring. 2008. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz, Weinheim.

[33] John McCarthy and Peter Wright. 2005. Putting ‘felt-life’ at the centre of human–computer interaction (HCI). *Cognition, technology & work*, 7, 4: 262-271. https://doi.org/10.1007/s10111-005-0011-y

[34] Tilo Mentler and Michael Herczeg. 2016. On the role of user experience in mission-or safety-critical systems. In *Mensch und Computer 2016–Workshopband*. Gesellschaft für Informatik e.V., Aachen, Germany, http://doi.org/10.18420/muc2016-ws01-0001

[35] Mark A. Musen, Blackford Middleton, and Robert A. Greenes. 2014. *Clinical decision-support systems.* In: Biomedical informatics, Springer, London, UK, 643-674. https://doi.org/10.1007/978-1-4471-4474-8_22

[36] Bala G. Nair, Eilon Gabel, Ira Hofer, Howard A. Schwid, and Maxime Cannesson. 2017. Intraoperative Clinical Decision Support for Anesthesia: A Narrative Review of Available Systems. *Anesthesia & Analgesia*, 124, 2: 603-617. 10.1213/ane.0000000000001636

[37] Julia Neily, Joseph M. DeRosier, Peter D. Mills, Michael J. Bishop, William B. Weeks, and James P. Bagian. 2007. Awareness and use of a cognitive aid for anesthesiology. *The Joint Commission Journal on Quality and Patient Safety*, 33, 8: 502-511. https://doi.org/10.1016/S1553-7250(07)33054-7

[38] Thierry Pelaccia, Jacques Tardif, Emmanuel Triby, and Bernard Charlin. 2011. An analysis of clinical reasoning through a recent and comprehensive approach: the dual-process theory. *Medical education online*, 16, 1: https://doi.org/10.3402/meo.v16i0.5890

[39] Meera Ramachandran, Joel S. Greenstein, Michael Mcevoy, and Matthew D. Mcevoy. 2014. Using a context-sensitive ranking method to organize reversible causes of cardiac arrest in a digital cognitive aid. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. SAGE Publications, Los Angeles, CA, USA, 788-792. https://doi.org/10.1177/1541931214581144

[40] Nicholas Riches, Maria Panagioti, Rahul Alam, Sudeh Cheraghi-Sohi, Stephen Campbell, Aneez Esmail, and Peter Bower. 2016. The effectiveness of electronic differential diagnoses (DDX) generators: a systematic review and meta-analysis. *PloS one*, 11, 3: https://doi.org/10.1371/journal.pone.0148991

[41] Paula Savioja, Marja Liinasuo, and Hanna Koskinen. 2014. User experience: does it matter in complex systems? *Cognition, Technology & Work*, 16, 4: 429-449. https://doi.org/10.1007/s10111-013-0271-x

[42] Daniel Schneiderbanger, Stephan Johannsen, Norbert Roewer, and Frank Schuster. 2014. Management of malignant hyperthermia: diagnosis and treatment. *Therapeutics and clinical risk management*, 10: 355-362. http://doi.org/10.2147/TCRM.S47632

[43] Kennon M. Sheldon, Andrew J. Elliot, Youngmee Kim, and Tim Kasser. 2001. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology*, 80, 2: 325-339. https://doi.org/10.1037/0022-3514.80.2.325

[44] Allan F. Simpao, Jonathan M. Tan, Arul M. Lingappan, Jorge A. Gálvez, Sherry E. Morgan, and Michael A. Krall. 2017. A systematic review of near real-time and point-of-care clinical decision support in anesthesia information management systems. *Journal of clinical monitoring and computing*, 31, 5: 885-894. https://doi.org/10.1007/s10877-016-9921-x

[45] Marjorie Stiegler, Jacques P. Neelankavil, Cecilia Canales, and A. Dhillon. 2011. Cognitive errors detected in anaesthesiology: a literature review and pilot study. *British journal of anaesthesia*, 108, 2: 229-235. https://doi.org/10.1093/bja/aer387

[46] Marjorie Stiegler and Keith J. Ruskin. 2012. Decision-making and safety in anesthesiology. *Current Opinion in Anesthesiology*, 25, 6: 724-729. http://doi.org/10.1097/ACO.0b013e328359307a

[47] Annette L. Valenta, Margaret M. Browning, Timothy E. Weddle, Greer Stevenson, Andrew D. Boyd, and Denise M. Hynes. 2010. Physician perceptions of clinical reminders. In *Proceedings of the 1st ACM International Health Informatics Symposium*. ACM, New York, NY, USA, 710-717. http://doi.acm.org/10.1145/1882992.1883101

[48] Jelle Van Dijk and Caroline Hummels. 2017. Designing for embodied being-in-the-world: Two cases, seven principles and one framework. In *Proceedings of the eleventh international conference on tangible, embedded, and embodied interaction*. ACM, New York, NY, USA, 47-56. http://doi.acm.org/10.1145/3024969.3025007

[49] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1-15. http://doi.acm.org/10.1145/3290605.3300831

[50] Robert L. Wears and Marc Berg. 2005. Computer technology and clinical work: still waiting for Godot. *Jama*, 293, 10: 1261-1263. http://doi.org/10.1001/jama.293.10.1261

[51] Leslie Wu, Jesse Cirimele, Kristen Leach, Stuart Card, Larry Chu, T Kyle Harrison, and Scott R Klemmer. 2014. Supporting crisis response with dynamic procedure aids. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, New York, NY, USA, 315-324. http://doi.acm.org/10.1145/2598510.2598565

[52] Qian Yang, John Zimmerman, Aaron Steinfeld, Lisa Carey, and James F. Antaki. 2016. Investigating the heart pump implant decision process: opportunities for decision support tools to help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 4477-4488. http://doi.acm.org/10.1145/2858036.2858373

[53] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1-11. http://doi.acm.org/10.1145/3290605.3300468

[54] Dimitrios Zikos and Nailya DeLellis. 2018. CDSS-RM: a clinical decision support system reference model. *BMC medical research methodology*, 18, 1: https://doi.org/10.1186/s12874-018-0587-6