DYNAMIC PROCEDURE AIDS: INCREASING ACCESS, ASSIMILATION, ACCEPTANCE, AND ATTENTION IN CRISIS RESPONSE AND HIGH-RELIABILITY DOMAINS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jesse Edward Cirimele

December 2013

This dissertation is online at: http://purl.stanford.edu/kr373hr8603

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Scott Klemmer, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Lawrence Chu

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Stuart Card

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

Many information-rich domains, from aviation to crisis response, require accurate responses under extreme time constraints. Paper checklists have been shown to improve important outcome measures. While these paper checklists are valuable, they are static, slow to access, and show both too much and too little information. Little is known about how to design checklists well, especially new forms of computer-aided checklists; experts produce most existing designs in an ad-hoc manner.

In response, we introduce the Dynamic Procedure aids approach. Dynamic Procedure aids address four key problems in checklist usage: ready access to the aids, rapid assimilation of their content, professional acceptance of their use in medical procedures, and the limited attention available to their users. This design solution arose through a multi-year participation in medical crisis response training featuring realistic team simulations. A study compared Dynamic Procedure aids, paper aids, and no aid conditions, finding that participants with Dynamic Procedure aids performed significantly better than with paper aids or no aid. This study introduces the Narrative Simulation paradigm for comparatively assessing expert procedural performance through a score-and-correct approach

Next, this thesis compares alternative checklist design proposals, distills effective design patterns, and shows that designing checklists with these features improves performance. A two-part experiment with medical participants was conducted in a laboratory with an eye-tracker. The first part compared time performance, eye-traces, and memory retention for five alternative checklist designs. From the results and design patterns, we distilled three key design principles to support rapid reading and instantiated them in a new design style. The second part compared the original designs to this redesigned style, called RapidRead. Applying these RapidRead principles reduced variance in response times, importantly minimizing the frequency of slow responses.

Acknowledgements

I thank Scott R. Klemmer for advising me through my Ph.D., always having time for me, and patiently helping me understand what it means to do good research; Stuart K. Card for his immense contributions to this research, and for always pushing for having theory and principles, and understanding the mechanism that explains our findings; Larry Chu for chairing my committee and reaching out from the medical school to invite us into a collaboration; Jeff Heer and Terry Winograd for great feedback on the project and for helping me truly learn HCI over my time at Stanford; Kyle Harrison for helping me understand crisis response by tirelessly explaining during simulation sessions, working with us to design a good aid, and helping us think through evaluations; Leslie Wu for co-leading with me on this project and helping to make it a success; Kristen Leach for collaborating with me in the papers and the systems of this dissertation; Justin Lee, Tonya Yu, Katherine Chen, and Kyle Barrett for their contributions during their summer internships; Jon Bassen and Lahiru Jayatilaka for contributing to the research during their rotations; Wendy Mackay, David Gaba, Steven Howard, Sara Goldhaber-Fiebert for the thoughtful discussions that contributed to this research; Abby King, Eric Hekler, Lauren Grieco, and the MILES team for letting me help their project and always making me feel appreciated. Tico Ballagas for his collaboration on my earlier projects and helping teach me the ropes of research; Jim Hollan, Virginia de Sa, and David Meyer for inspiring me in my Bachelor’s degree to engage in research, supported my learning, and recommended me for the PhD program; the participants of these studies; Maria Cirimele for always listening to my troubles, showing me how to focus on a single important project, and her constant companionship; my parents Ed and Carolyn and my brother Jason for their love and encouragement and for always being there for me; and my larger family for always supporting me. This research was sponsored in part by NSF POMI 2020 Grant No. CNS-0832820.

Table of Contents

List of Tables

List of Illustrations

Chapter 1

Introduction

Medicine, especially emergency medical procedures, is complex and time-paced, resulting in errors and adverse outcomes, about half of which have been estimated to be unnecessary. Simple checklists have been demonstrated to reduce errors, but checklists have themselves been criticized for slowing down medical procedures and competing for attention with the patient. In this thesis, we treat interaction with checklists as a problem in human-computer interaction with certain novel constraints. We analyze some of the weaknesses of the usual paper checklists and use the result to propose designs for replacing paper checklists with interactive displays that share information among team members, employ formatting to speed up use, are designed to be multitasked with patient care, and dynamically tune the information displayed to the task.

Figure 1 Dr. David Gaba running a simulated medical crisis.[1]

This dissertation introduces and evaluates interactive information systems for *complex perilous procedures* (CPP), such as those arising in surgery and hospital crisis responses. In contrast with the routine cognitive skill of many office tasks [Card et al. 1983], these procedures are at the edge of tractable complexity [Patterson 2007; Rochlin et al. 2005; Gawande 2009]. Errors are easy to make, yet the perilous environment is severely unforgiving of even small errors. Also, complex procedures, like medical crisis response and aircraft carrier work, require intricately coordinated multi-tasking, are team-based, and strongly time-paced [Patterson et al. 2002; Rochlin et al. 2005].

---

[1] Photo courtesy of http://www.flickr.com/photos/stanfordedtech

1.1. Complexity and Errors

Arguably, there is no complex perilous procedure domain with more impact than surgery and emergency medical care. The number of surgeries performed globally each year is about 234 million and rising [Haynes et al. 2009]. However, the complexity of surgery and related medical crisis care leads to a higher level of adverse outcomes than necessary [Gawande 2009]. The influential “Harvard Study” looked at deaths in hospitals and estimated there are between 44,000 and 98,000 deaths that resulted from preventable harm per year [Brennan et al. 1991; Leape et al. 1991]. More recent estimates put that number closer to 400,000 [James 2013]. Other studies estimate about half of adverse outcomes are preventable [Kohn et al. 2000; Davis et al. 2002; Vincent et al. 2000; Neale et al. 2001; Dekker 2011]. To avoid harm, many tasks must be executed almost perfectly by highly-skilled teams working tightly together under significant time pressure. One study counted 178 tasks per day for the average patient in an Intensive Care Unit [Donchin et al. 1995]: each task puts the patient at risk. Beyond the need for almost faultless execution, the vast number of conditions and remedies increases the complexity. The WHO international disease classification system lists 13,600 diagnoses, 6000 drugs, and 4000 medical and surgical procedures [Center For Disease Control 2005].

To manage this complexity, doctors have begun to adopt risk-management techniques from aviation, such as training in simulation environments, crew resource management, and the use of checklists [Gaba et al. 1994; Gaba et al. 2001]. Aviation, like nuclear power and space flight, is a CPP domain where checklists have been studied for decades [Boorman 2001; Burian et al. 2005; Degani and Wiener 1990;

Gawande 2009; Harrison et al. 2006; Haynes et al. 2009; Ziewacz et al. 2011]. Applied to medicine, checklists do, in fact, reduce errors for both simulated medical crisis response [Arriaga et al. 2013; Harrison et al. 2006; Ziewacz et al. 2011] and routine tasks such as pre-surgery setup [Makary et al. 2006] or inserting central lines [Pronovost et al. 2006].

1.2. Checklists

Using even simple checklists can substantially reduce adverse events [Pronovost et al. 2006]. Without checklists, physicians skipped at least one step while putting in a central line (a catheter used to administer fluids) about a third of the time [ibid]. Using a checklist reduced ten-day infection rates from 11% to 0% in his first study. Introducing checklists into Michigan hospitals decreased infection rates by 66% and saved about $175 million, and more than 1500 lives, in the first 18 months. When extended to hospitals of different types and countries, introducing checklists saw major complications from surgery drop 36% and deaths 47% [Gawande 2009]. As McConnell writes, checklists are an important “vessel of safety culture” [McConnell et al. 2012]. These benefits generalize to procedurally organized knowledge, often called cognitive aids [Chu and Fuller 2011]. Anesthesia care teams improve their performance with increased use of cognitive aids [Harrison et al. 2006].

Though medical checklists have drawn inspiration from aviation, there are important differences between the domains, especially in team composition and work. In aviation, the physical ergonomics are static and highly regulated. Aircrews sit in cockpits where controls and displays are co-designed and co-located [Hutchins 1995]. By contrast, operating rooms (ORs) have sensors, information displays, and interaction points spread throughout the environment [Mentis et al. 2012; Sarcevic et al. 2010]. Cockpit crews work in small teams of two or three, and typically have similar backgrounds. Hospital crisis care teams may comprise surgeons, anesthesiologists, pharmacists, nurses, technicians, and other specialists, arranged *around* the patient, each with their own cultures, roles, and equipment. Not only must staff work under time pressure, risk, and uncertainty; they must also cope with the coordination and communication complexity inherent in team-based crisis care [Hunziker et al. 2011]. These complexities lead to breakdowns in effective crisis care: missed steps, timing errors, lack of a shared mental model, and poor resource management.

Checklists have the potential to mitigate and recover from these breakdowns, but they must be carefully implemented or they could exacerbate potential problems. For example, finding and searching checklists can induce additional time, attentional demand, and complexity [McConnell et al. 2012]. This procedural interference—and cultural skepticism—has slowed checklists' adoption by medical teams [Gawande 2009; Winters et al. 2009]. As Verdaasdonk et al. [2009] put it: "Time governs willingness and compliance in the use of checklists." In psychological terms, *perceived* time may be the more salient variable. Even Gawande, one of checklists' foremost promoters, noted the usability failure of his first attempt to make a viable checklist [Gawande 2009]. Furthermore, given that medical checklists are designed as cognitive aids, it is ironic that checklist deployments sometimes give the impression of an externally imposed barrier or disruption, ignorant of skill, wisdom, and context. Thomassen, et al. [2010] note that "Despite the increasing use of checklists in healthcare worldwide, few studies have explored personnel experiences in using this new tool." Those that have (*e.g.*, [Fourcade et al. 2012]) find barriers such as the checklist slowing or otherwise disrupting the procedure. Though current checklists often provide benefits, their costs have impeded changes in medical practice. This leads us to believe that a stronger benefit:cost ratio could tip the scales.

1.3. Approach

Our approach to checklists was to use participant observation to induct a set of core problems to be solved. From these we developed a set of core design concepts to address the core problems. These guided a set of prototyping studies that explored possible designs. Finally, we tested our resulting design with experiments based on a technique we call Narrative Simulation.

Over the first 16 months working with doctors, we observed more than 50 hours of simulated crisis scenarios at a state-of-the-art, high-fidelity, medical simulator on campus. We observed medical teams responding to both operating room crises and cardiac arrests. Furthermore, we engaged in a number of design critique sessions with participants in these simulated operating rooms. Based on these sessions and a participatory design process with collaborating doctors, we created and analyzed more than 50 static design variations for Dynamic Procedure aids, as well as two functional web prototypes. We identified four key design concepts: Ready access, rapid assimilation, professional acceptance, and limited attention. This observation and design work created the first version of Dynamic Procedure aids (chapter 3).

Our goals were to evaluate these aids on medical doctors using a method that was fast, efficient, and had easily comparable results. We developed the Narrative

Simulation approach to accomplish this (chapter 4). Narrative Simulation—inspired by the MegaCode video training materials [ACLS-Algorithms 2012]—presents scenarios in a linear fashion, no matter how the participants responds. The participants saw a slide-based presentation that automatically advances to tell the patient story. For example they may initially learn that the patient is a 64-year old male with a certain blood pressure and heart rate. Later, the scenario presents that the heart rate has changed. The scenario slide then asks the participant how they will respond. The participant's response is recorded and assessed for accuracy. The system then presents the canonically accurate response, describes the action taken and continues the story. This linearity and synchronization enables comparison across participants and conditions at each step. In the Narrative Simulation evaluations, when participants used the Dynamic Procedure aids they answered more questions correctly than when they used paper aids or didn't use any aids (chapter 4).

To get a deeper understanding of aid use, our second study looked at information finding speed and eye-trace data on different aid designs (chapter 5). We choose speed because there is a limited amount of time doctors can switch attention away from their primary task to attend to a cognitive aid. If information assimilation cannot be completed in a single glance, it will be ignored or remembered and continued on another cycle [Salvucci and Taatgen 2008]. In addition, medical tasks in emergency medicine are also time-paced: there is limited time before a deadline by which they must be done. The design concept of fitting a checklist or cognitive aid step into a multi-tasking cycle we call a *step-at-a-glance* user interface.

It is not only speed that is important here. Predictability is perhaps even more important. It is better to have a low variance by being consistently fast than to be extremely fast most of the time but sometimes extremely slow. Therefore, step-at-a-glance aids must quickly and reliably tell the doctor what she needs to know when she needs to know it.

We introduce a set of techniques for increasing the speed of searching for information on demand in cognitive aids called *RapidRead* (chapter 5). The compact object-action language frees space on the display and visual chunks grouped for faster search by providing a short verbal handle. Visual chunks constrain visual search and this effect is amplified by the dynamic adjustment of focus+context.

Health professionals (medical doctors and EMTs) were given cognitive aids and asked questions related to Advanced Cardiac Life Support. Each question was asked for five different aid designs and required participants to visual find the answer. Response times were measured and eye-gaze recorded using an eye-tracker. Dynamic aids were shown to have faster and more consistent response times. Using variations on paper and Dynamic aids, RapidRead design principles were verified and improved.

1.4. Contributions of this work

This thesis makes contributions in three areas:

- We introduce Dynamic Procedure aids and highlight their four key design concepts that came out of our design work: ready access, rapid assimilation, professional acceptance, and limited attention. Two studies found that

Dynamic Procedure aids perform better than paper aid styles in both Narrative Simulations and information finding tasks.

- The RapidRead approach introduces the object-action writing pattern, organized into information patches and revealed/navigated through a focus+context layout. These guidelines help by reducing information finding time and eye-traces are more structured.
- We developed Narrative Simulation, a scenario-driven evaluation technique. It works well for domains where system usage can't be isolated from context/scenario of use. Potential applications include testing of ubicomp or high-risk interfaces where you can't quickly, easily, cheaply test in the actual setting [Shami et al. 2005].

Our results suggest that Dynamic Procedure aids could also improve on checklists in other domains, such as space flight, aviation, and complex machinery.

In general, CPP domains are good a good match for checklists, and thus our techniques, because errors have high cost, tasks have many steps, and time pressure is high. In the future we can continue this work by evaluating in high-fidelity simulations, personalizing digital aids, and studying interaction with dynamic aids.

Chapter 2

Related Work

Chapter 1 introduced complex perilous procedures, and highlighted some of the major findings from the motivational work on checklists. This chapter dives more deeply into three areas: medical errors and teamwork, human factors around multi-tasking and teamwork, and interfaces designed to improve multi-tasking and teamwork.

2.1. Medical errors and teamwork

Doctors work at the edge of tractable complexity [Gawande 2009; Patterson 2007; Rochlin et al. 2005]. There are an overwhelming 13,000 issues a patient can be treated for, 6,000 different drugs, and 4,000 medical and surgical procedures [Gawande 2009]. In addition, doctors must also have nearly flawless execution of the their treatment procedures [Gawande 2009]. While it is tempting to look at the low accident and error rates other high-reliability organizations [Rochlin et al. 2005] and just translate their successes over to medicine, this hasn't proved to be a simple task [Patterson 2007]. For example, the culture of continual training and learning on aircraft carriers has developed organically in part because there is nearly one-hundred percent turn-over every forty months [Rochlin et al. 2005]. A solution that works well in this idiosyncratic environment may not translate well to a health care domain without understanding the human factors that underlie the success.

2.1.1. Medical errors

A review of studies that looked at preventable medical errors conservatively puts them as the 8th-leading cause of death in the US [Kohn et al. 2000]. What errors are we solving with checklists and what errors are not being solved?

Looking at ten themes from a detailed analysis of five medical adverse events (medical treatments with harmful or undesired complications) [Patterson et al. 2002] we can split them into two groups. Issues from Patterson et al. that will likely not be helped by crisis checklists include: goal conflicts with safety, reduced resources and expertise, poor coordination across service "silos", poor observability of patient status, and degraded ability to detect a problem and recover. These issues have to do with available personal and team resources and the need to be ready prior to the crisis.

The issues from Patterson et al. that crisis checklists may help solve include: complexity, deviation from nominal workflow, missed side effects of change, poor dynamic task re-allocation, and poor hand-off briefing. These are issues where having the right information or reminder at hand can be the difference between correct and incorrect action.

Specifically, we can look at the sub-items cited under complexity: distributed work and interconnections between roles, high workload, time pressure, complex medical regimen, non-standard use of medication, and new treatment [Patterson et al. 2002]. These issues are quite similar to the motivating issues of CPP domains, and the prominence of these complexity-based problems shows that complexity-reducers like checklists should help.

2.1.2. Medical teamwork and coordination

Teamwork and coordination warrant consideration because medical teams are multi-disciplinary, and patient safety as well as clinical performance depend on them [Kolbe et al. 2011]. When teams transition into crisis they must also quickly adapt their coordination techniques, such as spending more time on information management to keep team members apprised of each other's actions [Burtscher et al. 2011]. Medical simulations have been used teach and test teamwork and coordination in a realistic manner [Gaba et al. 1994; Gaba et al. 2001; Kharasch et al. 2011; Bong et al. 2010]. Teams that do well in simulated crisis scenarios spend more time on coordinating activities, like thinking out loud, whereas teams that do poorly spend more time working independently and don't efficiently use available resources [Manser et al. 2009].

Medical teams have looked to aviation to help improve teamwork and coordination by adopting resource management techniques [Gaba 2011a; Gaba et al. 1994], also called non-technical skills [Nestel et al. 2011; Gaba 2011b]. In order to treat a patient, doctors must not only know how to operate their tools, but they must also know task management skills like planning and preparing; team work skills like confirming roles and responsibilities, identifying a leader, and sharing information; situation awareness skills like vigilance in monitoring and anticipating problems; and decision making skills like balancing risks and re-evaluating the situation [Flin et al. 2010].

2.2. Human Factors

2.2.1. Multi-tasking: Attention and Interruptions

Doctors multi-task during crisis response, including diagnosing the patient, treating the patient, coordinating with the team, and continuing to monitor the patient's vitals. Multi-tasking is characterized by a serial execution of multiple streams of thought. For example, you can be cooking a couple of different dishes simultaneously, but you can only attend to one at a time [Salvucci and Taatgen 2008]. Multi-tasking can allow a set of tasks to be performed in less time than it would take to do each task serially [Frisch et al. 2012], but this isn't without a heavy cost. People respond more slowly and make more errors immediately after a task switch [Monsell 2003], and nobody is as good at multi-tasking as they think they are [Sanbonmatsu et al. 2013].

Other challenges doctors face during crisis are that time pressure increases effort, as does keeping multiple things in the memory at once [Kahneman 2011]. Another risk of errors emerges when one is fully mentally engaged with a difficult problem, because they can be so focused that they completely miss visual and auditory information in their environment [Kahneman 2011].

Finally, task interruptions happen constantly as new information is revealed during the crisis, but interruptions can lead to forgetting what they were previously doing [Cutrell et al. 2000] and similar prospective memory errors [Dismukes and Nowinski 2007]. This means each time you switch mid-tasks while multi-tasking you not only lose time, but you also run the risk of forgetting something.

2.2.2. Teamwork and Coordination

Acting in concert as a team is not an easy task. Much prior work has tried to understand how teams coordinate by looking at situation awareness, shared mental models, communication, and grounding.

A shared mental model refers to a shared understanding of a situation between two or more people. Then, when a decision needs to be made, the team can act using shared well-structured knowledge, without spending a lot of time aligning their understanding [Mathieu et al. 2000; Cannon-Bowers et al. 1993]. More recently, researchers believe that shared situation awareness is constructed through social interactions on the fly, rather than through primarily pre-existing knowledge [Heath et al. 2002].

This is similar to how people construct shared understanding in conversations [Clark and Brennan 1991]. This grounding is a conversational back and forth so that both individuals in the conversation know that the other person understands them [Clark and Brennan 1991]. This is easier, requiring less time and complexity in words, if people in the conversation share a visual reference [Gergle and Clark 2011].

2.3. Designing for multi-tasking and low attention

For inspiration in how to deal with the medical domain, we can look to prior HCI research on interfaces for multi-tasking and group coordination.

2.3.1. Peripheral displays and multi-tasking

Peripheral and ambient displays [Vogel and Balakrishnan 2004] provide information to support a task without requiring consistent attention [Matthews et al. 2009; Maglio and Campbell 2000]. Peripheral displays have been used for monitoring secondary tasks, such as keeping good posture while sitting at a computer [Jafarinaimi et al. 2005], as well as for supporting primary tasks, such as supporting balanced face-to-face discussion [Bachour and Kaplan 2009]. One of the earliest examples of a peripheral display was Natalie Jeremijenko's "Dangling String", in which a hanging string would wiggle proportionally to local Ethernet traffic [Brown and Weiser 1996]. This poetic presentation portrays otherwise opaque information at a glance, and acts as an inspiration to the kind of interface that we want to build.

Peripheral displays are, by definition, designed for multi-tasking and thus have two major considerations that single-focus systems don't have: Awareness and Distraction [Matthews et al. 2009]. Awareness issues deal with the ease of extracting information from the display. By this definition, increasing awareness means making displays faster to understand and easier to monitor. For example, using visual afterglow effects to highlight critical changes that may have happened while the user wasn't paying attention might help people monitor peripheral displays [Baudisch et al. 2006]. The other primary goal for peripheral designs is to reduce distraction from the primary task, for example by minimizing motion [Maglio and Campbell 2000].

Another technique for handling multi-tasking to include modeling user workload in order to only interrupt the primary task when it will have the lowest cost [Adamczyk and Bailey 2004; Bailey and Iqbal 2008; Iqbal and Bailey 2005].

In the medical domain, doctors must maintain awareness of the patient's status, progress in treatment, and team members' actions. Vitals displays are a prime example of a peripheral display. Machines display vitals with enough history so that if doctors need to look away for short periods of time they don't miss critical changes. They minimize motion by using a trace line that sweeps across the screen instead of a scrolling display [Seagull et al. 2001]. This means that, in contrast to a scrolling display, once a data point has been placed on the screen it doesn't move until the screen has cycled back around and written over it.

2.3.2. Designing Checklists

Checklists and related tools have been built in many domains. A few prior papers have compared alternative presentation styles for tasks such as programming [Brandt et al. 2010], and furniture assembly [Agrawala et al. 2003]. However, this empirical literature is sparse and does not address rapid information acquisition tasks in externally-paced domains like medical crisis response.

Work on paper checklists in aviation has focused on the importance of good typography in making checklists easy and fast to read [Degani 1992]. In medicine, design advice, in the form of a checklist for making checklists [Gawande 2013], provides guidelines for distilled from experience, such as using "fewer than 10 items per pause point".

Chapter 3

OBSERVATIONS AND DESIGN CONCEPTS

In our 16 months working with doctors, we observed more than 50 hours of simulated crisis scenarios at a state-of-the-art, high-fidelity, medical simulator on campus. We observed medical teams responding to both operating room crises and cardiac arrests. From behind a one-way mirror, we saw dozens of medical residents work with confederate actors to handle complex and unexpected patient crises. We sat in on the post-simulation debriefs organized by the medical teaching staff, as they taught the principles of crisis resource management [Gaba et al. 2001]. Furthermore, we engaged in a number of design critique sessions with participants in these simulated operating rooms.

3.1. Participants and Process

High-fidelity medical simulation (see Figure 2) offers a unique opportunity to investigate crisis response, without endangering live patients and without posing privacy concerns. These simulations were created to provide a safe, realistic setting for medical education and doctor re-certification [Gaba et al. 2001]. They place one or more students in an operating room with a confederate crew of nurses and doctors. Behind the scenes, the simulation team uses theater-like mixing boards along with computers to remotely control the patient mannequin responses and directs the in-room confederate crew in order to realize the scenario for the students. The patient mannequin has enough functionality to make it all feel real: it has a pulse, it breathes, it can take IV fluids, it's eyes can dilate, it's forehead can sweat, and it can even make small seizure-like movements.

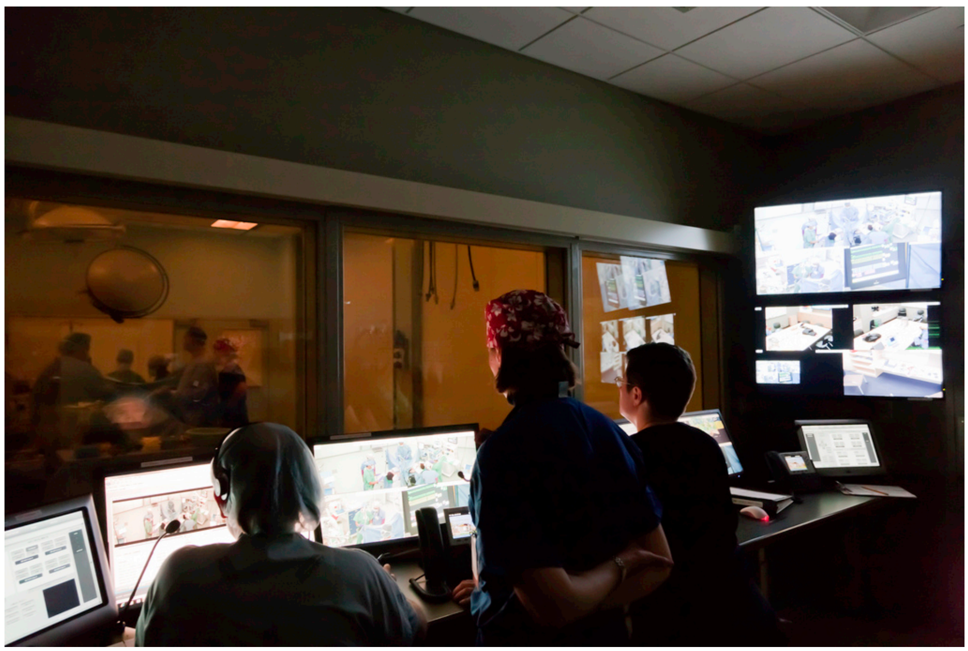


Figure 2 High-fidelity crisis simulations are run in realistic settings with theater production.[2]

Our observation focused on the practice of operating room (OR) anesthesiology. In the hospital setting we observed that OR anesthesiologists are responsible for managing emergent events during perioperative patient care.

[2] Photo courtesy of http://www.flickr.com/photos/stanfordedtech

Anesthesiologists are trained to recognize and respond to medical emergencies and take on the role of team leader once this occurs. Like pilots, anesthesiologists prepare for the beginning of surgery ("take-off"), keep an eye on the controls, the end ("landing"), and have been characterized as having "hours of boredom punctuated by moments of terror" [Gaba et al. 1994; Gaba 2007; Rehmann et al. 1983]. Both fields have different checklists and protocols for routine care versus crisis care.

Figure 3 (left) an early paper prototype with embedded WHO checklist. (right) Another early prototype where we tried out timeline and patient monitors. These were eventually discarded to simplify.

We created more than 50 different dynamic checklist prototypes at various fidelities (see Figure 3), sat in on actual surgery, and reviewed video recordings of simulated crises with medical faculty to walk through user errors and opportunities for software system interventions. To understand the interaction demands, an interactive Web application was developed using HTML and JavaScript. This prototype used WebSockets to synchronize tablet and large-screen displays and was deployed in two high-fidelity operating room simulations. Initial prototypes addressed general surgery. Later prototypes concentrated on a task domain of cardiac arrest treatment in a hospital setting because of its ubiquity and importance. This includes a set of about eight cardiovascular diagnoses that result in life-threatening crisis situations. For example, Bradycardia is a potentially fatal arrhythmia that causes the heart to beat dangerously slow; Tachycardia is dangerously fast; and Asystole is defined by the absence of electrical activity in the heart. The medical term for this domain is Advanced Cardiac Life Support (ACLS) [Neumar et al. 2010].

3.2. Key Design Concepts

This section describes the four problem areas that emerged from our participatory design (Table 1): 1) *ready access* (making the aids themselves more rapidly and reliably accessible to the team), 2) *rapid assimilation* (decreasing the time to find and assimilate information from the aid), 3) *professional acceptance* (increasing team acceptance of the aid), and 4) *limited attention* (improving the ability to multitask with the aid). For each area, we identified a key design concept formulated to shift the aid to reduce the cost or increase the benefit for the medical staff. We also provide the concrete instantiation of that concept in the Dynamic Procedure Aid System and our design rationale for why it should help.

Table 1: The four key issues; their induced design shifts; and proposed solution components.

NEED	KEY CONCEPT	DESIGN INSTANTIATION(S)	HOW IT ADDRESSES PROBLEM
1. Ready Access: Hard to find; Hard to share	**Shared Display:** Make aids visible to team through large-screen display. DESIGN SHIFT: Paper ➔ Multiple shared displays	Mirror display and interaction across multiple large-screens and tablets	Provides shared context, facilitates finding checklist, provides more detail
2. Rapid Assimilation: Too slow; Hard to multitask with patient care	**Steps-at-a-Glance:** Procedure step processable in one multitasking cycle. Focus on what to do now in abbreviated context. Simplify Display. Speed reading and search. DESIGN SHIFT: Text ➔ Object/State + Information mapping	Reformulation of step to be findable and readable in small bursts. Object/Action, compressible checklist language. Progressive aid protocols.	Faster read, skim, search due to: - reduction in number of words - stereotyped syntax - Information mapping Processable in small time units for multitasking
3. Professional Acceptance: Mixed acceptance leading to less use	**Resources-at a-Glance:** Reframe checklists as part of a larger, resource management system. DESIGN SHIFT: Checklist ➔ Resource Management	Rapid access to team names, supplies, calculators, reference Allow aid to transition from routine to crisis, display additional resources	Provides incentive to use system, familiarizes and habituates practitioners
4. Limited Attention: Narrow, scarce attention under stress	**Attention Aids:** Direct interface focus dynamically DESIGN SHIFT: Attention regulator ➔ Attention Aid Focus+Context	Automated drug timers and attentional prompts	Cognitive aid serves as attentional aid

3.2.1. Ready Access

Problem: The Invisible Paper Aid.

One fascinating and unexpected observation was that doctors responding to a crisis would often start using a paper cognitive aid until they found an item on the aid that needed to be done. Then, they would put the checklist aid down on some flat surface, where almost invariably it would be covered by something else and never picked up again. Other times, doctors would hold the binder containing the aids in one hand, without a convenient place to position the binder so that it was visible and yet accessible. Furthermore, doctors were inconsistent in their use. Given identical scenarios, some doctors never picked up the aid, others looked at it once, and others made personal and/or public use of its information. Consequently, the aid's useful information was often invisible, hidden physically, or held by only one team member.

We also observed work practices and mental models. For example, one doctor informed another of an important change in patient vitals, the other doctor failed to hear, but this was not obvious. As a result, neither realized they held different mental models of the situation. Unsurprisingly, coordinated mental models correlate with improved team performance in both aviation [Mathieu et al. 2000] and medicine [Manser et al. 2009].

Key Concept: Shared Displays.

We hypothesize that a large, shared display can mitigate both forgotten aids and misaligned beliefs. It can provide a consistent physical location, legible from most locations, supporting common ground [Clark and Brennan 1991; Clark 1996]—the achievement of "a shared understanding of what is being discussed in a conversation with multiple participants" [Birnholtz et al. 2010]. By providing shared visual referents to the procedure, its state, and the resources involved, the grounding process may be shortened. For example, if the display indicates which drugs were administered, a query about them might be answered with a quick gesture or might not have to be asked at all. In other words, we hypothesize a shared display can make cooperation more *implicit* and less *explicit* [Entin and Serfaty 1999], increasing speed and reducing errors.

Our early software prototypes ran on a single large-screen display mounted on a wall. However, any single location had blind spots for someone because to face the patient would mean someone had to turn their back to the wall (See Figure 4). Subsequent prototypes added a second display so that everyone had a clear view. These displays can be permanently mounted in an OR, or brought in on “crash carts” wheeled in during emergency codes. One benefit of a single, shared display is the clarity of what everyone can see—in contrast with personal displays where it’s less clear what individuals can see [Wallace et al. 2009]. Synchronizing the two displays retains most of the grounding clarity that a shared display provides. Furthermore, a synchronized view enables input to be driven by an individual, such as a nurse with a tablet. Our process found nurses to be a valuable intervention point because of a professional inclination to process adherence and functional role in organization and support. Doctors could also give verbal commands to nurses for controlling the display.

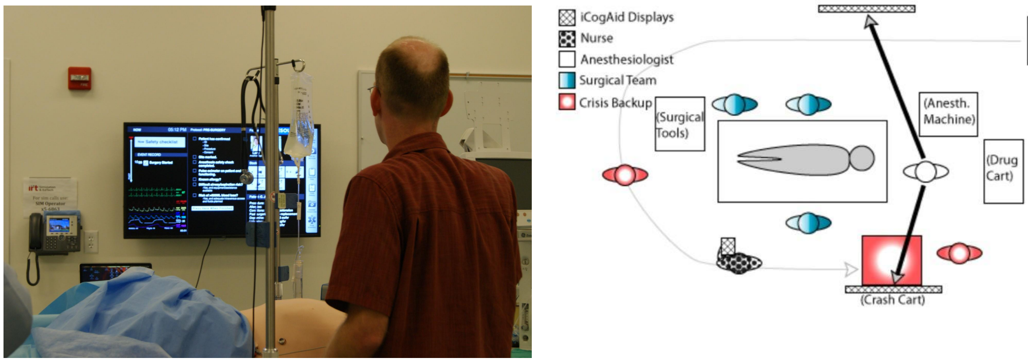
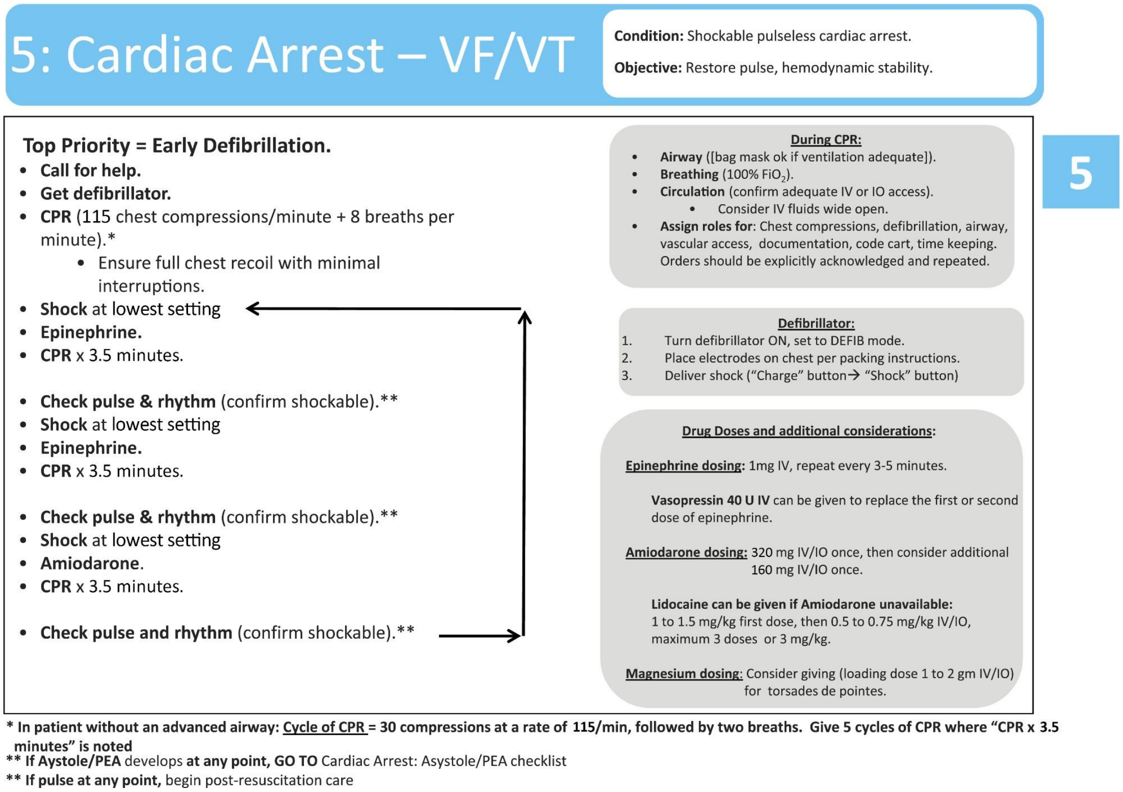


Figure 4: Doctor refers to digital aids on a large-screen display (left); Example OR layout (right)

3.2.2. Rapid Assimilation

Problem: Too Much, Much Too Slow

Checklist critics claim they are slow to use and consequently compete with time and attention needed for the patient [Kendell and Barthram 1998; Winters et al. 2009; Verdaasdonk et al. 2009]. Checklists have differential benefits for distinguish rare procedures from common ones. For rare medical events that a team has never experienced, checklists provide new or poorly recalled information. Here, checklists aids must be easy to understand. By contrast, for common events, checklists cover routine and familiar material and serve as a reminder to not skip steps or make assumptions too quickly. Here, checklists should be easy to skim, and remind effectively. In between, checklists are used to look-up or confirm a particular fact, such as a drug dosage. In all cases, to support rapidly shifting visual attention, steps must also be fast to find and to re-find if the reader looks away to attend to something else.



5: Cardiac Arrest – VF/VT

Condition: Shockable pulseless cardiac arrest.

Objective: Restore pulse, hemodynamic stability.

5

Top Priority = Early Defibrillation.

- **Call for help.**
- **Get defibrillator.**
- **CPR** (115 chest compressions/minute + 8 breaths per minute).*
    - Ensure full chest recoil with minimal interruptions.
- **Shock** at lowest setting
- **Epinephrine.**
- **CPR** x 3.5 minutes.

- **Check pulse & rhythm** (confirm shockable).**
- **Shock** at lowest setting
- **Epinephrine.**
- **CPR** x 3.5 minutes.

- **Check pulse & rhythm** (confirm shockable).**
- **Shock** at lowest setting
- **Amiodarone**.
- **CPR** x 3.5 minutes.

- **Check pulse and rhythm** (confirm shockable).**

During CPR:

- **Airway** ([bag mask ok if ventilation adequate]).
- **Breathing** (100% FiO_2).
- **Circulation** (confirm adequate IV or IO access).
    - Consider IV fluids wide open.
- **Assign roles for**: Chest compressions, defibrillation, airway, vascular access, documentation, code cart, time keeping. Orders should be explicitly acknowledged and repeated.

Defibrillator:

1. Turn defibrillator ON, set to DEFIB mode.
2. Place electrodes on chest per packing instructions.
3. Deliver shock ("Charge" button→ "Shock" button)

Drug Doses and additional considerations:

Epinephrine dosing: 1mg IV, repeat every 3-5 minutes.

Vasopressin 40 U IV can be given to replace the first or second dose of epinephrine.

Amiodarone dosing: 320 mg IV/IO once, then consider additional 160 mg IV/IO once.

Lidocaine can be given if Amiodarone unavailable: 1 to 1.5 mg/kg first dose, then 0.5 to 0.75 mg/kg IV/IO, maximum 3 doses or 3 mg/kg.

Magnesium dosing: Consider giving (loading dose 1 to 2 gm IV/IO) for torsades de pointes.

* **In patient without an advanced airway: Cycle of CPR = 30 compressions at a rate of 115/min, followed by two breaths. Give 5 cycles of CPR where "CPR x 3.5 minutes" is noted**

** **If Aystole/PEA** develops **at any point, GO TO** Cardiac Arrest: Asystole/PEA checklist

** **If pulse at any point,** begin post-resuscitation care

Figure 5: Paper checklists provide valuable information. This checklist [Ziewacz 2011] exemplifies how static information presentation can be hard to skim during crisis response.

Key Concept: Steps At a Glance.

A useful way of designing for multitasking [Salvucci and Taatgen 2008; Brumby et al. 2007] is to estimate a typical time interval during which the dominant task can be neglected and to design steps of the secondary task so that they can be completed in this turn length [Green 1999]. We introduce the *step-at-a-glance* concept that information artifacts should be designed so that steps can be assimilated in one glance. This chunking speeds use and facilitates attentional shifts when needed. Our participatory design led to three techniques that reduce the time of assimilating a step.

Balance simplicity and amount of information. Our early designs included nearly every piece of information that participants suggested, and consequently suffered from feature clutter (see Figure 3 right). This led to a display where in principle everything was available but in practice little was findable. Technical, information-rich domains often face this tension. Our challenge was exacerbated by the wall-scale form factor, which requires clear legibility at a distance. A lesson we learned repeatedly was that the scalpel can fruitfully be applied to interfaces.

Focus on current context. In reviewing prototypes, doctors strongly preferred a clear and simple representation of the current context, even when that required sacrificing useful but more peripheral information. Like turn-by-turn map directions, the whole screen can be focused on the current step, simultaneously increasing relevant information and reducing cognitive load [Jeung et al. 1997]. While paper is restricted to a static display, software can emphasize currently needed information (see Figure 6), such as a specific treatment protocol. Information that has already been used, is not yet needed, or provides additional explanation for the curious can be minimized by default and expanded if necessary. This approach expands the focus+context layout strategy [Card et al. 1999; Bedersen 2000] to procedural documents.

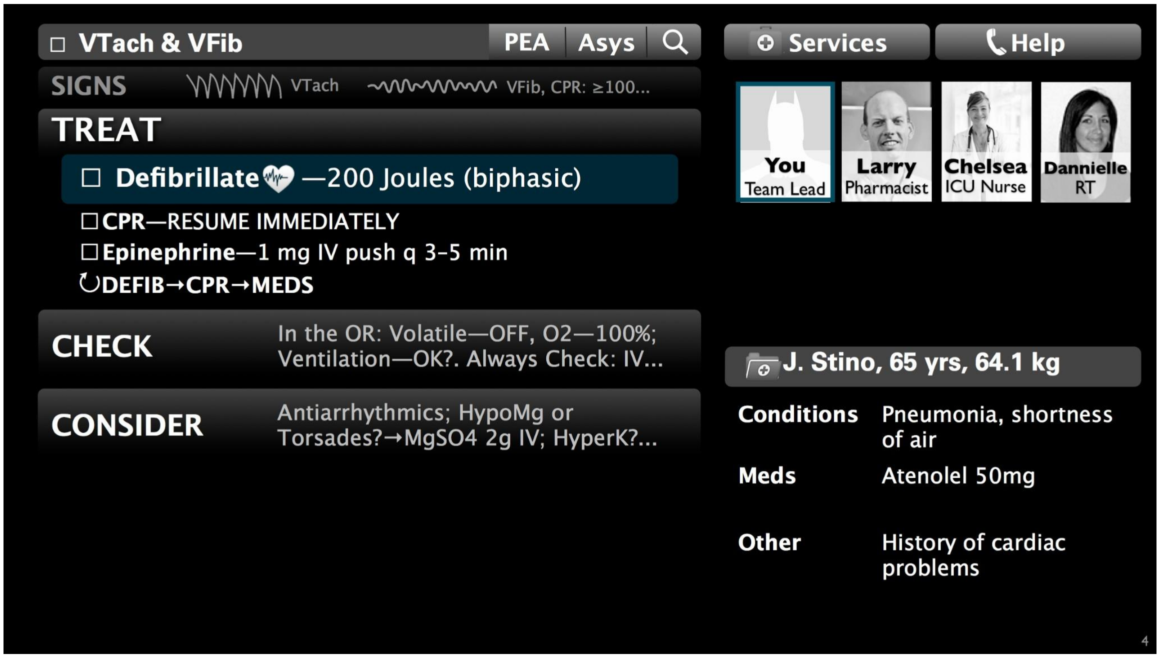


Figure 6: Dynamic Procedure Aid for Ventricular Tachycardia & Ventricular Fibrillation

Object/Action checklist language. Early medical crisis checklists [Ziewacz et al. 2011] were presented as full sentences with comparatively little visual structure (*e.g.*, Figure 5). This is different from early aviation checklists where utilize the highly constrained information structure and let the visual design can carry more of the information load and improve usability [Burian 2004; Burian 2006]. Chu's aids leveraged this with richer visual presentation [Chu and Fuller 2011]. Our work continued in this vein, extracting the basic procedural structure from written descriptions and representing it graphically when appropriate. Increasing visual structure and shortening the text speeds reading and improves scanning (see chapter 5). We have designed a stylized language for re-expressing medical procedures in an object/action compressed language. This language, loosely inspired by configuration checklists for aircraft [Burian 2004], reduces the number of words in a checklist, in some cases by half. Whenever possible, each step begins with an object followed by an action or state setting to be achieved for the object. For example, the steps

Increase FiO_2 to 100%
Verify ischemia with 12 lead EKG if possible

could be re-expressed as

FiO_2: ↑100%
Ischemia: Verify *(Use 12-lead EKG)*

We further exploit the structure by listing the object to the left, bold facing it, and giving it larger type, creating a consistent information mapping [Horn 1990] from content to visual form. We furthermore expand the steps of the procedure (see Figure 7) when they are at the point of execution to make available additional subsidiary information. Collectively, these treatments are designed to increase speed for the several types of procedure reading: direct reading, skimming, and searching.

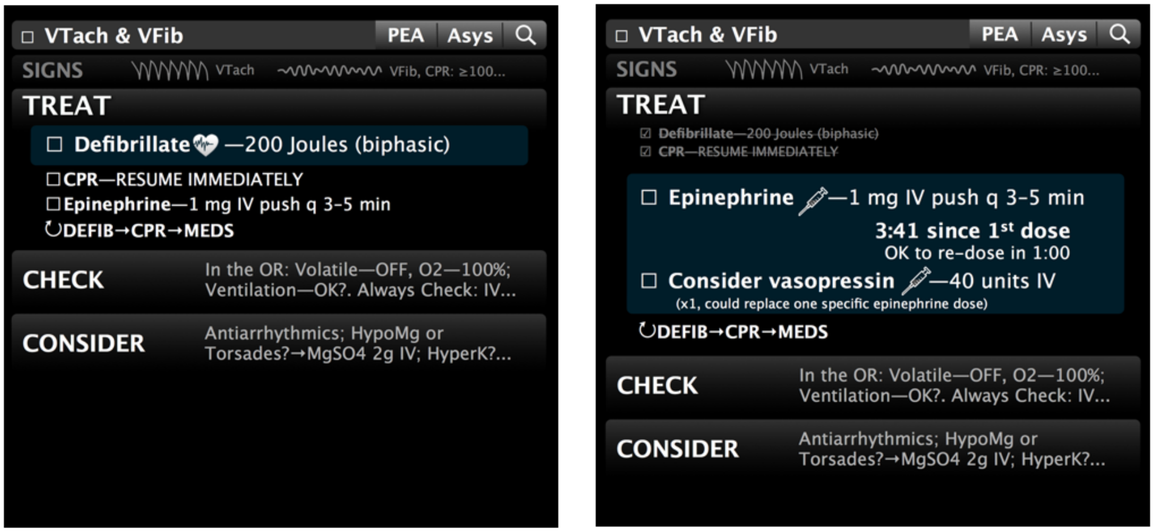


Figure 7: Compressed language combined with variable disclosure: selecting an element in the overview (left) reveals additional details (right).

3.2.3. Professional Acceptance

Problem: Bridging the Gap.

As we have described, current checklist aids often improve outcomes, yet are underused because some perceive an unfavorable cost:benefit ratio or an unwelcome and unwise restriction on professional autonomy.

Key Concept: Resource at a Glance

According to literature reviews by [Degani and Wiener 1990; Degani and Wiener 1993; Verdaasdonk et al. 2009], checklists should serve the following functions:

- a defense strategy to prevent human errors
- a memory aid to enhance task performance
- standardization of the tasks to facilitate team coordination

• a means to create and maintain a safety culture

• support for quality control by management, government and inspectors

Highly-skilled professionals rarely welcome the oversight implied in the later items of this list, even if this standardization on average improves outcomes. Even in aviation, where checklist use is standardized, too many checklists reduce compliance [Hales and Pronovost 2006]. At the same time, professionals in many fields seek better, timely information. In one simulated crisis we observed, an anesthesia resident pulled out his smartphone to search the Internet for information about a competing diagnosis (thyroid storm). Because the form factor of the information was ill-suited for the device and task, he spent about 5 minutes out of a 20 to 25 minute crisis reading his device. We see this as evidence that bite-sized, contextually-relevant information is a critical need. Therefore, we propose adding to this list another function:

• rapid access to task-relevant information mid-crisis

That is to say, we propose generalizing *checklists* into *procedure aids*.

To address these perceived and actual cost:benefit problems we expand the benefits, reduce the usage costs, and emphasize the cognitive aid role over the bureaucratic oversight role. Our work reframes the checklist aid concept to feature them as the centerpiece of an integrated resource view. For example, at a large hospital, team members commonly don’t know the names of everyone else on the team, especially when their medical attire, including surgical cap and facemask, obscure much of their head and face. This often yields open-loop communication such as “we need to get the crash cart” rather than closed-loop communication (*e.g.*, “Jon can you call for the crash cart”, Jon—“yes I will call for the crash cart”). The Dynamic Procedure aids shared screen shows pictures and names of people in the room along with information about those on their way to help (see Figure 8). This simple cognitive aid makes the social space visible and, potentially, the communication more precise.

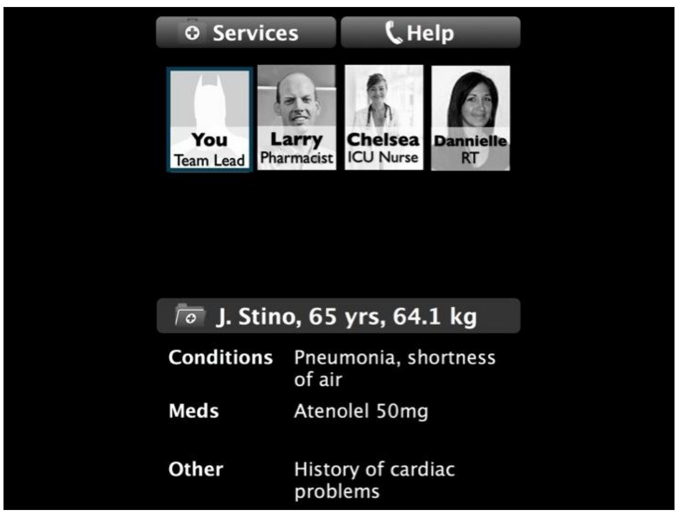


Figure 8: Integrating additional resources, like patient and team information, helps make the dynamic aid a “one-stop shop”, encouraging usage.

Our prototypes also explored dashboards showing inventories of blood, medicine, and other supplies available, the expected time to availability of laboratory tests, patient records needed for the procedure, patient identification and procedure site, and the plan of the procedure, names and roles of the operating team, and images useful for the procedure. It can embed medical calculators already initialized to the patient's weight and other parameters. Each time a team member would like to gather resource information, they look to the same screen. By providing an integrated, glanceable view of multiple, commonly-referenced resources, we hope to lower the activation energy for acquiring information, facilitate serendipitous reminding, and create the habit of more frequently consulting these resources.

Figure 9: An Anaesthesiologist monitors a standard patient vitals display.

3.2.4. Limited Attention

Problem: Paced, Multi-surface, Multi-user Attention.

Attention is a major limiting factor during crisis response [Takahashi et al. 2011]. Multiple co-located people work across multiple surfaces on a network of interdependent, important tasks. For example, anesthesiologists may split visual attention between a vitals display (see Figure 9), the patient, and a drug vial they are preparing, while simultaneously ensuring that other staff continue high-quality CPR. Medical personnel must orient and attend cognitively, physically and socially. This physically-distributed attention [Srinivasan et al. 2009] differs from desktop [Horvitz et al. 2003] and mobile [Iqbal and Bailey 2010] attentional patterns, and complicates the design of software for crisis teams.

Medical doctors aren't alone in resisting lockstep adherence. For example, in aviation, electronic checklists for routine operation (pre-flight checklists) sometimes mandate step-by-step affirmation. These draconian systems have been poorly received because even pilots don't usually check off every item, the so-called READ-DO method [Gawande 2009]. During routine operation, they mostly use the READ-CONFIRM method of performing several items from memory, then consulting the checklist to see if they missed anything. This chunking of multiple operations saves time and hassle. And during crisis response, the required speed of crisis response makes step-by-step affirmation unworkable.

People naturally modulate their care in response to challenge and risk [Bergen et al. 2013]. For pilots, the extreme hazards of transoceanic flight engender greater diligence, and they often employ the more cautious READ-DO approach. We hypothesize that enabling this flexibility increases adoption.

Which doesn't mean people always make the right call when left to their own devices, so we must also make the adherence path encouraged and fast. The reader role in medicine [Burden et al. 2012], the WHO surgical time-out [Makary et al. 2006], and the READ-DO or READ-CONFIRM practice in aviation [Gawande 2009] exemplify the attention regulator approach. In this approach, an agent (reader in medicine, co-pilot in aviation) blocks, or regulates, progress until some action is taken.

Administering recurring drugs provides a frequent and important example. We observed that frenetic pacing and multiple responsibilities caused medical teams to sometimes forget to miss the time to re-dose, or forget about a prior dose and re-dose too often. Some operating rooms rely completely on memory, others have a nurse track dosages on a clipboard or whiteboard [Aronsky et al. 2008]. Precisely timed attention to multiple activities is difficult for people, but easy for software. And timers can serve as a clear, high-value draw that in turn engenders broader use.

We initially explored audio alarms, because they are more agnostic to physical orientation. However, operating rooms are extremely noisy [Healey et al. 2006]: even during routine operation, rock music combines with device alerts, social chat, and work-related discussion. For example, anesthesiologists may be listening to the surgeon while asking a nurse to call for an arterial blood gas, peripherally keeping an ear out for the O_2 saturation, but ignoring a false alarm from a different machine. We also learned during our observations that currently there are no regulations on how medical alarms should behave, so the alarm tone, volume, and frequency are as varied as the device manufacturers. Crises make matters worse: though social chatter dissipates and music is turned off, the number and frequency of genuine and false alarms increases dramatically, as does the speed and volume of communication. Consequently, “demanding” attention through an audio alert is often fruitless and possibly detrimental. However, the medical professionals we talked to told us that, somewhat like pilots, they trained to cycle rapidly through the dashboard displays they are responsible for, and a visual alert can be ready for them when they do.

Key Concept: Attention Aid

Given these complexities, our design shifted from checklists as attention *regulators* to checklists as attention *aids*. To help medical teams maintain state, Dynamic Procedure aids provide context-specific drug timers and alternate diagnoses to consider. The timers embed dosage and countdown information at the relevant step of the cognitive aid, concentrating relevant information where it's needed (see Figure 7). Suggestions such as "consider _____ aid" flag medically similar diagnoses and diagnoses the current condition may evolve into. These suggestions lower the cost of switching to another aid. Suggestions also seek to discourage fixation on initial diagnosis, a common issue under duress [Burian and D 2006; Gaba et al. 2001; Burian 2006]. Like the timers, Dynamic Procedure aids place these suggestions within the aid at the relevant action step to facilitate their use.

Chapter 4

Experiment 1: Test of Dynamic Procedure aids for checklist guided procedures

How well is our design for dynamic aids (our further development of the checklist idea) likely to work in practice? In this chapter, we test our design concept by comparing it against current best practices as well as some alternatives. The most ecologically valid test would be to use our aids in real emergency medical situations. However, this alternative presents at least three problems. First, it is unethical to put patients at risk with untested technology. Second, it could be expensive and take a long time to run. And third, since each real situation would have unique factors and a unique medical team, the situation is not well-controlled. Even the use of medical simulators does not resolve this problem. Although we would gain more experimental control, experimental tests would be too slow and expensive for iterative engineering design. Sanders [1984] has eloquently discussed this tension between ecological validity and the experimental control necessary to draw valid conclusions.

Our solution is to develop Narrative Simulation to investigate the hypotheses that Dynamic Procedure aids would be easier to use, be used more frequently, and would help doctors perform more effectively than their paper counterparts. Narrative

Simulation—inspired by the MegaCode video training materials [ACLS-Algorithms 2012]—presents timed scenario slides in a linear fashion regardless of how the participant responds. The participant sees a slide-based presentation which automatically advances to tell the patient story. For example, they may initially learn that the patient is a 64-year old male with a certain blood pressure and heart rate. Later, the scenario presents that the patient's heart rhythm has changed. The scenario slide then asks the participant how they as the doctor will respond. The participant's response is recorded and assessed for accuracy. The system then presents the actual action taken in the scenario and continues the story. This linearity and synchronization enables comparison across participants and conditions at each step.

The goal of Narrative Simulation was to create a fast and inexpensive evaluation that would allow us to test and compare the presentation-action aspects of cognitive aids. Participants are asked to verbalize proper procedure under attentional stress and time limits. These Narrative Simulation scenarios were designed to place the participants in the role of team-lead for cardiac arrest crisis. Participants were asked to act as the team leader and make decisions and judgments about treatment. This allowed us to test aspects of cognitive aid usage and verify their merit before investing in full medical teams and expensive simulation setup. Although not as realistic as a high-fidelity medical simulation, the Narrative Simulation explicitly incorporated narrative elements that emulate medical teamwork. For example, scenarios explicitly involve virtual team members that report vitals and ask for next steps.

4.1. Method

4.1.1. Participants

Thirty-seven people (28 M.D.'s and 9 medical students) were recruited from our university to participate in this one-hour study: 20 female and 17 male, all trained medical personnel. Participants included 2 fellows, 27 residents, and 9 medical students. Participant specialties were distributed as follows: Internal Medicine (8), Anesthesia (7), Emergency Medicine (7), Undecided (3), Surgery (2), Dermatology (1), Radiology (1), and Urology (1). All were trained in Advanced Cardiac Life Support (ACLS), which requires re-certification every 2 years. The distribution of recertified participants was: two years ago (4), one year ago (13), in the current year (16), and not yet certified (4). In this hospital, residents are responsible for running the cardiac arrest response teams. These skills have a moderate number of opportunities for practice: 2-4 cardiac arrests (codes) per month, in which residents may be involved in only one of these every couple of months. This work is under human subjects protocol IRB-25138.

4.1.2. Materials.

Pre-Study Survey. The pre-study survey asked participants for their (expected) graduation year from medical school, medical specialty, date of first ACLS certification, and date of their most recent ACLS certification. Participants were counterbalanced based on their amount of ACLS training (group 1: zero or one certifications, group 2: two or more).

Scenario Design and Slide Simulators. This study created narrative encapsulations of authentic medical scenarios. Scenarios were designed to test participants' medical knowledge and crisis management. These medical scenarios were adapted from the MegaCode online training videos [ACLS-Algorithms 2012] and modified by our medical collaborators.

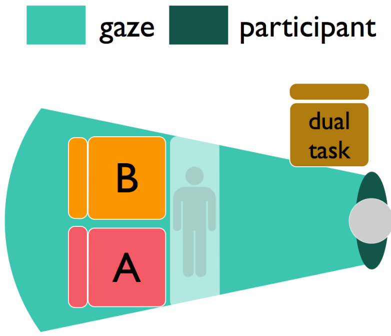
In this simulator, scenario slides advance automatically every 5 seconds and reveal information about the patient and how the crisis progresses. During the scenario between 20 and 30 questions appear and participants are given 10 seconds to answer them verbally. Questions that were not answered within the limit were counted as incorrect. Regardless of how the user chose to act, the scenarios remain on a predetermined narrative path.

In order to determine correct/incorrect scoring, we took three steps. First, we generated a rubric with the help of our medical doctor collaborator who regularly teaches and evaluates the crisis response material we were testing in the medical school. Second, two graders did one third of the participants together to align their expectations, and split the other two thirds equally. Finally, graders re-evaluated ambiguous answers with the doctor who helped create the rubric.

Paper Cognitive Aids. In this condition, participants were provided with the paper ACLS checklist aids developed by Gawande et al. [Ziewacz et al. 2011]. These aids were chosen because they had been previously validated in the literature [Ziewacz et al. 2011] and shown to support crisis teams responding to ACLS scenarios in high-fidelity simulations. We printed the paper cognitive aids out on 8.5" x 11" paper and laminated them so they would be sturdy and easy to handle. They were placed on a table nearby, a common practice.

Dynamic Procedure Aids. In the Dynamic condition, the slide simulator was augmented with a second screen showing the Dynamic cognitive aid. To synchronize with the Narrative Simulation, the Dynamic Procedure aid interface was presented on pre-timed slides. These slides synchronized with the scenario slides, advancing as if a nurse or reader were controlling the interface. The medical content in this condition was substantively equivalent to the paper condition. Dynamic Procedure aid content was based on existing paper aids, but the content focus dynamically changed based on the scenario context. This approach allowed us to evaluate the proposed interface design quickly with minimal implementation complexity.

Experimental Setting and Apparatus. The experimental room was configured with an empty patient chair, a laptop displaying the scenario, and a secondary task apparatus (see Figure 10). The scenario screen presented the simulation narrative and prompted the participant with questions. Audio and video were recorded. In the dynamic condition, an external display showed the Dynamic Procedure aid. One laptop ran the checklist; a second ran the simulator questions; a third ran the secondary task. In the paper condition, participants were provided with laminated paper aids on a table.



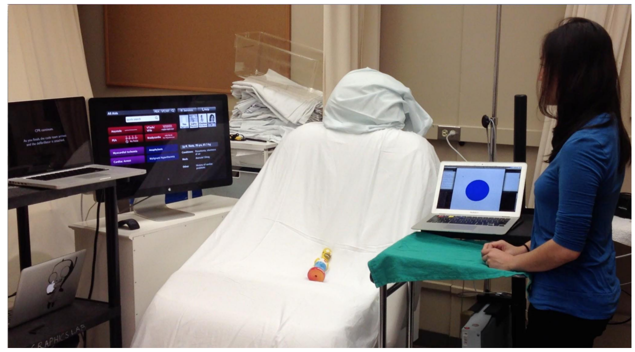


Figure 10: Overhead view of experimental setup with scenario [A] and displayed aid [B] (left). Participant uses dynamic aid while responding to questions (right), with color task visible and adjacent.

Secondary Task. To simulate the additional cognitive load and multi-tasking required in many crises, participants were required to attend to a secondary task. On a separate screen, a filled circle randomly changed colors from gray to red, yellow, or blue approximately 50 times during each scenario. Participants had 10 seconds to press the color-labeled keyboard key corresponding to the correct color, changing it back to gray. The required multi-tasking created an additional load on the participant's attention, since the participant had to physically turn to monitor the secondary task display.

One of the concerns when using a secondary task is that how the participants choose to focus their attention and effort becomes an uncontrollable confounding variable when analyzing performance on the primary task. If one participant performs well on the primary task and poorly on the secondary task, while another participant performs well on the secondary task and poorly on the primary task, it becomes quite difficult to compare them on their performance on the primary task. To mitigate this issue, the difficulty of the color task was chosen such that participants would uniformly do well. It is then possible to compare performance on the primary task.

4.1.3. Procedure

Experimental Sequence. The experiment comprised the following steps: consent form, pre-study survey, training, scenarios (*Pneumonia*, *Syncope*, *Unresponsive*), followed by post-scenario surveys, post-study survey, and final debriefing. Total participant time for the study was 1 hour. Simulation runs were audio and video recorded. All participants were exposed to three screen-based simulations, always in the same order. Nurses and other doctors were implicitly present in the scenario design.

Training (10 mins). Participants were guided through a ten-minute training period to familiarize themselves with the simulation slides, secondary task, paper cognitive aids, and the dynamic checklists. Participants ran through two abbreviated versions of ACLS slide simulations, first with paper cognitive aids and next with a synchronized dynamic checklist.

User Scenarios (3 x 8 mins). The following outlines the sequence of medical conditions presented in each scenario.

Male, 65, Pneumonia*: Bradycardia, Asystole, Ventricular Fibrillation (25 questions)*

Male, 65, Syncope*: Unstable Supraventricular Tachycardia, Ventricular Fibrillation (25 questions)*

Female, 78, Unresponsive*: Ventricular Fibrillation, Asystole, Ventricular Tachycardia (24 questions)*

Conditions. There were three conditions: the participants could receive Dynamic Procedure aids, they could receive the paper aids, or they received no aid for the scenario. Participants saw each condition once. Participant condition order was counterbalanced using a latin square design. Participants in the conditions with aids were told, "In this condition you will be given access to an Aid. It will be located here." They were not explicitly told they had to use the aids.

Post-Scenario Self-Assessment (3 x 1 min). After each scenario, participants filled out a survey on their perceived performance for the secondary task and medical scenario response. They were asked to respond to these three questions:

> *1. Out of fifty color changes, how many times do you feel like you selected the incorrect color or missed one entirely?*
>
> *2. Out of thirty questions, how many questions do you feel like you missed?*
>
> *3. If you used a cognitive aid/checklist, how much do you feel it changed your score on the questions? (give a number positive or negative)*

Post-Study Survey and Final Debriefing (10 mins). After the three scenarios, participants filled out a final survey, including demographic information (gender), open response questions about ACLS expertise, and about their checklist experience. A grading rubric was used to score valid and invalid responses to scenario questions such as "What is the next important step?" or "What is this [EKG] rhythm?" Partial credit was given depending on the timing of the response, and specificity (reduced credit for incorrect dosage but appropriate drug or defibrillation). Graders were research assistants familiar with the scenarios and the appropriate ACLS response. Two different researchers reviewed the grades to ensure consistency and accuracy. All materials required to replicate the experiment, including the secondary task, surveys, scenarios, paper aids, Dynamic Procedure aids, and experimental protocols are available online for download at https://hci.st/dpAid/study-2013

4.1.4. Statistical Analysis and Data Cleaning

All scores are reported as the percentage of correct trials. Results were compared using fixed effects modeling[3], a variety of linear regression. Our analysis was done in R using the "lm" function. Unlike the t-test and similar to the ANOVA, linear regression is able to account for the probability of multiple pair-wise tests being true at the same time. In addition, these models have two benefits over a repeated-measures ANOVA. First, a linear model with only fixed effects can handle unbalanced or missing data, and is otherwise equivalent to a multivariate ANOVA used for repeated measures analysis. Fixed effects linear models are strictly more powerful than an ANOVA. Second, random effects can be added to account for factors such as participant and scenario differences that in practice cannot be exhaustively sampled [Baayen et al. 2008]. In this dissertation we primarily use fixed-effects regression models. Each reported result comprises three pieces: first, per-condition averages; second, the effect-size β, indicating the slope difference reported by the mixed effects

---

[3] http://en.wikipedia.org/wiki/Fixed_effects_model

model; and third, the key statistic and *p*-value. Note that β is slightly different than simply subtracting the condition averages because β incorporates the model's estimate of the underlying variation in the random and fixed effects.

Data Cleaning. Twenty-nine of 37 starting participants had data that we could analyze across all scenarios: six had at least one of their scenarios removed due to synchronization issues; and two were exposed to incorrect conditions. In the *Pneumonia* scenario, we removed questions 16 to 24 from the analysis after discovering that a software bug that caused Dynamic Procedure aids to get stuck in the wrong state on those questions for all participants. We report the results after this data cleaning.

4.2. Results

Aid type. Dynamic Procedure aids reduced medical procedure errors. Participants using Dynamic Procedure aids responded correctly significantly more often than unaided participants did (79.6% vs 69.1% correct; $\beta = 9.46$, $t(82) = 3.3$, $p<.01$), but those using paper aids were not statistically different than unaided (70.0% vs 69.1%; $\beta = .30$, $t(82) = .104$, $p = .92$) (see Figure 11). Moreover, more use of the Dynamic Procedure aid correlated with fewer errors (Adj $R^2 = 0.28$, $F(4,82) = 8.013$, $p < .001$).

Looking only at the first experimental scenario creates a between-subjects comparison that avoids the risk of priming or fatigue affecting the data. When we use data only from the first scenario, the effect of Dynamic Procedure aids becomes even stronger: those using Dynamic Procedure aids responded correctly significantly more often than unaided participants (80.0% vs. 63.6% correct; β=16.4, $t(26) = 4.3$, $p <$

0.01). Again, there was not a significant performance difference between paper aids and no aids (67.6% vs. 63.6%; $\beta = 3.95$, $t(26) = .974$, $p = .34$).

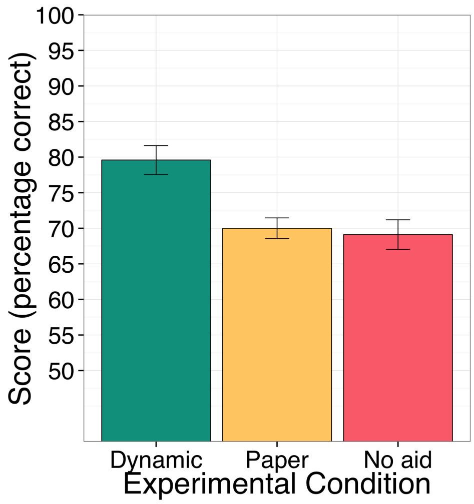


Figure 11: Participants using Dynamic Procedure aids responded correctly significantly more often than those using paper aids or no aid.

Significant factors and interaction Effects. To determine what factors were important in predicting scores, we compared several different models. In R, to compare two "lm" models, we used the "ANOVA" function on pairs of model outputs. If the ANOVA was significant, that indicated the two models were different. By incrementally adding in factors and interaction effects and testing for significance, we found that scenario, experience level, and experimental condition were all important for predicting score. In addition, the interaction between scenario and experience level, the interaction between experience level and experimental condition, and the interaction between experimental condition and scenario were all not significant. This indicates that there were no significant interaction effects for these factors.

Scenario difficulty. Scenarios varied in difficulty, as measured by error rate. The *Pneumonia* and *Syncope* scenarios did not differ significantly ($\beta = -1.2$, $t(82) = -.042$, $p = .67$), but the *Unresponsive* scenario was easier than the *Pneumonia* scenario ($\beta = 9.1$, $t(82) = 3.17$, $p < .01$).

Experience. As might be expected, advanced medical personnel (residents and fellows) had more correct trials than medical students when controlling for condition and scenario (74% vs. 67%) ($\beta = 8.3$, $t(80) = 2.81$, $p <.01$).

Secondary task. Across all scenarios, participants successfully responded to 92% of colors. There was a learning effect: response rates improved as scenarios progressed (88%, 93%, 97%). There was a marginally significant effect of condition on the total missed responses on the secondary color task (85 dynamic, 88 none, 115 paper, $\chi^2(2, n=30)=5.7$, $p=0.06$).

Perceived Utility of Aids. Participants, according to the post-test survey, perceived both paper aids and digital dynamic aids as beneficial. However, participants perceived a larger increase in score when using Dynamic Procedural aids (15.3%) than paper aids (4.4%) ($t = -4.52$, $df = 56.0$, $p < .001$).

4.3. Experimental Discussion

4.3.1. Exploring the benefits of Dynamic aids

Overall, participants using Dynamic Procedure aids responded correctly significantly more often in the simulated medical procedure than those using paper checklists or no aids at all (79.6% to 70.0% and 69.1%). Dynamic Procedure aids focus on four problem areas of medical checklists: ready access, rapid assimilation, professional acceptance, and limited attention. We discuss observations related to each.

Ready Access: paper aids can be tough to find, easy to lose, and inconvenient to hold. Dynamic aids sought to mitigate this problem by giving participants a shared display that always showed a relevant aid and resources. The study found that indeed participants used Dynamic aids more than paper ones (mean 22.9 vs 18.1 times per participant. $t=-2.2$, $df=54$, p-value $< .05$).

Rapid Assimilation: Current aids are slow to read and search, and this diverts important attentional resources away from the patient. Dynamic aids sought to mitigate this problem through its step-at-a-glance design pattern of cuing attention to the current step, and placing all step-relevant information in that one location. To achieve glanceability, we re-expressed the content of aids in an object-action stylized language that places objects and actions at a consistent location.

The secondary task simulates the doctor's multiple attentional demands. This dual-task methodology converts attentional load into errors [Martin 2007]. Consequently, the Dynamic aids' lower error rate suggests that the step-at-a-glance pattern was effective in reducing attentional load.

Professional Acceptance: Dynamic aids feature a prominent digital display that integrates multiple resources. This prominent presentation appears to have been successful: participants estimated that Dynamic Procedure aids improved their score by 15.3%; paper aids by 4.4%. This difference is significant ($t = -4.52$, $df = 56.0$, $p < .001$). It is important that cognitive aids both improve performance and are perceived to improve performance. Combined with the improvements in rapid assimilation and integration with resource management, we believe the techniques in this interface can help earn acceptance for cognitive aids. One risk of studying people who volunteer for research studies on cognitive aids is that it's possible there was a self-selection bias among participants. It will be important for future work to assess the perceived efficacy in the broader medical community.

Limited Attention: Crises have multifarious activities competing for scarce cognitive resources. Attentional overload more acutely affects people with less training and practice, because each task requires more conscious effort and attention [Ericsson and Lehmann 1996]. Consequently, a change in the novice/expert performance spread often indicates a change in the attentional bandwidth required. Improving newcomers' performance is especially important because they commit more errors [Phillips and Barker 2010]. In this study, unsurprisingly, doctors had a higher accuracy rate than medical students (74.5% v. 67.0%, see Figure 12). However, medical students' performance increased far more with the Dynamic aid (21% for students, 7.5% for doctors). This suggests that Dynamic aids are more attentionally efficient, providing users with more headroom for the intrinsic demands of the tasks.

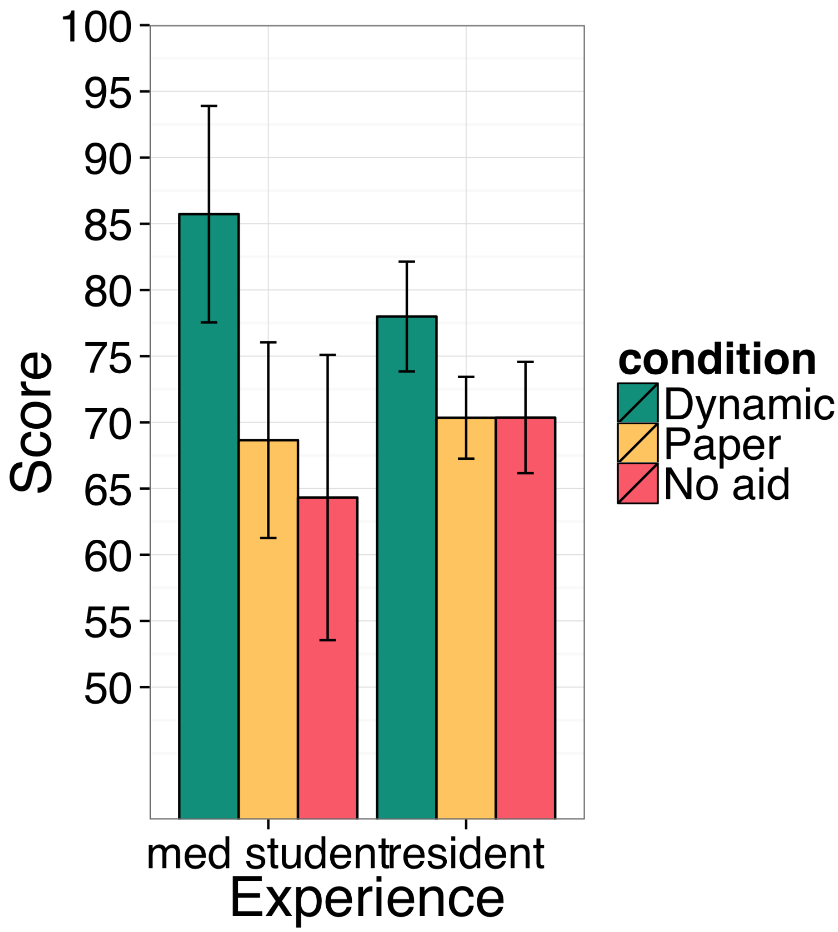


Figure 12: While overall residents outperformed medical students, students received significantly larger benefit from using Dynamic aids.

Note that medical students with Dynamic aids seemed to outperform more experienced residents who also were provided Dynamic aids. One explanation is that students trusted aids more and thus received more benefit. In contrast, residents may have trusted their experience over aids.

4.3.2. (When) do paper aids help?

Notably, this study found no significant advantage of paper aids compared to no aid. By contrast, several prior studies *have* found increases in team performance metrics and protocol adherence [Ziewacz et al. 2011; Harrison et al. 2006; Arriaga et al. 2013]. We see three likely factors for this difference: usage, teams, and training.

First, this study did not force participants to use any aids, paper or Dynamic, since the study was specifically interested in measuring voluntary use. In the Dynamic condition, everyone referred to the aid at least ten times. By contrast, in the paper condition, five participants used the aid fewer than ten times. When participants elected not to use the aid, they were essentially placing themselves in a no-aid scenario. A post-hoc t-test comparing infrequent (< 10 uses) and frequent (10 or more) paper aid usage did not find a significant impact of aid usage on score. Future work should explicitly assess the impact of mandated versus voluntary usage, as well as understand the role of time pressure in checklist usage and assimilation.

The second difference is that prior studies have evaluated the impact of cognitive aids on medical teams, which includes the coordination benefits that aids may provide. This study measured individual performance—when by definition there is no team coordination to be done.

The third difference is that prior studies probably offered more training with the particular cognitive aids studied. In this study, participants received about two minutes of training with each aid style. In the debriefing, participants often reported that the factor most inhibiting their use of paper aids was lack of familiarity. Many had different aids that they had practice with and therefore preferred. Given these constraints, it is particularly striking that participants were able to use the digital aid well with such minimal training. Aggregating the results of this study and the prior literature suggests that paper aids are valuable when used, but that underuse may minimize their practical impact, and that digital aids may provide a smoother and more effective adoption path.

4.3.3. Advantages and Limitations of Narrative Simulation

Narrative Simulation is a new style of evaluation for scenario driven interfaces and knowledge, similar in scope to other peripheral display evaluation techniques [Shami et al. 2005]. Other potential evaluation contexts used in medicine are real surgery and high fidelity simulations. For our purposes, real surgery is not an option. Real crisis situations are relatively rare (2 - 4 a month in our moderately sized research hospital) so they are hard to schedule for. In addition, and perhaps more importantly, life and death situations are not an appropriate initial testing ground for novel interface concepts. Our discussion here will therefore focus on Narrative Simulation and high-fidelity simulation.

In contrast to high-fidelity simulation, the goals for Narrative Simulation are not high realism or perfect understanding of a full system in context, but rather the rapid evaluation of novel components to enable rapid design iterations. A strength of Narrative Simulation is its relative speed. A single moderator can run participants on multiple scenarios in an hour and use of the tool can be observed. By contrast, high fidelity simulations require eight to twelve supporting doctors and staff in order to test two people at a time and require up to an hour for each scenario and debrief. In using Narrative Simulation, we follow standard engineering practice of iteratively designing and testing progressively more integrated prototypes in progressively more realistic application environments [Dym et al. 2013]. The development of systems passes through a sequence of testing methods. In the aerospace industry, for example, this sort of testing has been formalized in terms of “Technology Readiness Levels” [Mankins 1995; Layton 2003]. Our use of Narrative Simulation would correspond roughly to Technology Readiness Level 3, Analytical and experimental critical function and/or characteristic proof-of-concept. The use of high fidelity medical simulators for testing a mature prototype system might correspond to Technology Readiness Level 6, System/subsystem model or prototype demonstrated in a relevant environment. Use in a real crisis might correspond to Technology Readiness Level 7, System prototype demonstration in an operational environment.

Beyond development speed and expense, Narrative Simulation helps address the tension between experimental control and ecological validity. Narrative Simulation may, in fact, produce *more relevant information* in a *more controlled environment*. Thus, Narrative Simulation is not an alternative to high fidelity simulation but a complement that enables rapid, experimentally controlled studies to be paired with ecologically valid studies enabled by the high fidelity simulator, a point that has been eloquently argued by Sanders [1984]. For example, Narrative Simulation allows easy comparison between participants. All participants are asked exactly the same questions at exactly the same time in the simulation after seeing exactly the same information. By contrast, high-fidelity simulations are an intricate dance between doctors behind the scenes controlling patient vitals and giving instructions to the doctors and nurses in the room who are playing supporting roles, while the doctors in the hot seat are reacting and making treatment decisions. Much like real life, no two high-fidelity simulations are exactly the same after ten minutes.

Narrative Simulations have other clear trade-offs in comparison to high-fidelity simulation. High-fidelity simulation is used to study coordination and communication issues in dyads [Manser et al. 2009] and at our research hospital it is used for training and maintenance for larger teams [Gaba et al. 2001; Gaba et al. 1994]. Our evaluations were on individuals on their role within a simulated team environment, which means that issues of coordination and communication are difficult to evaluate. In principle, Narrative Simulations could be designed to be run on dyads, as high-fidelity simulations are. This would require specific understanding of the dynamics that were being simulated, but provides an interesting avenue for future work.

Finally, participants' performance in high-fidelity simulations is often scored by actions that relate to improved patient outcome [Manser et al. 2009; Harrison et al. 2006; Ziewacz et al. 2011]. Narrative Simulation scores performance by answering questions about best practice treatment steps. This emulates a common practice in medical education, where medical students are questioned verbally in clinical settings to assess knowledge and familiarity with procedures [Wear et al. 2005]. Low-fidelity simulation techniques such as screen-based, interactive medical simulation software have been shown to improve, and relate to, performance in higher-fidelity simulation environments [Nyssen et al. 2002].

Thus, while this Narrative Simulation study found significant benefits of Dynamic Procedure aids, future work should evaluate them using methods that compliment Narrative Simulation and broaden our understanding of when they are useful.

Chapter 5

Experiment 2: Testing RapidRead Design with Eye-Tracking

In addition to their role in guiding the execution of medical procedures, cognitive aids also contain information, such as drug doses, machine settings, and diagnostic or other information that the medical team may need to extract quickly. In this chapter, we refine the design of dynamic aids and codify a set of heuristic rules for generating them. We then test their performance for information abstraction tasks using response time and eye movement analysis. Two experiments with medical participants were conducted in a laboratory equipped with an eye-tracker. The first experiment compared time performance, eye-traces, and memory retention for five alternative checklist designs. From the results, we distilled three key design principles to support rapid reading and instantiated them in a new design style. A follow-up experiment tested retention and compared the original designs to this redesigned style. Applying these principles–which we refer to as RapidRead–reduced variance in response times, importantly, minimizing the frequency of slow responses.

5.1. Making Checklists Fast

In routine, self-paced tasks slow checklists might be an inconvenience. In externally-paced tasks like driving and surgery, the longer you are diverted from your primary task, the slower your response time will be during critical events possibly leading to severe consequences [Monk et al. 2008; Horrey and Wickens 2007; Wickens and McCarley 2007].

Medical Crisis checklists have additional challenges. Human bodies are complex and treatment can't be as linear as checklists for engineered processes [Gaba et al. 1994], there is nearly always a large treatment team with a mix of specialties [ibid], interruptions are common [Chisholm et al. 2000; Healey et al. 2006], and checklists aren't yet enforced in most hospitals [Gawande 2009]. This means that medical crisis checklist usage is not a prescribed Read-Do as it is in routine cases or aviation, but rather it is a Do-Confirm or just a lookup tool for infrequently performed steps.

How fast do crisis checklists need to be? There hasn't been a formal study of the impact of checklist speed on patient outcome, so we have to look to related domains. In aviation, there has been several accidents where slow, difficult to read checklists contributed to the problem [Degani 1992] and flight crew speed in accessing, reading, comprehending, and executing procedures impacts saftey during emergency or abnormal conditions [de Ree 1991]. In driving safely literature, extensive studies on driver's gaze, accidents, and in car interfaces have shown that off-road gaze time correlates with more accidents [Green 2002; Green 1999; Horrey and Wickens 2007]. Furthermore, the longer that reading a checklist diverts from the primary task, the higher the chance of prospective memory errors [Wickens and McCarley 2007; Dismukes and Nowinski 2007]. Finally, the ability to rapidly acquire information from external resources increases people's usage [Kalnikaité and Whittaker 2007; Verdaasdonk et al. 2009; Sparrow et al. 2011], which is important for reducing errors and ensuring up-to-date responses. The work in these other domains all support the conclusion that checklists should be designed to be fast in order to minimally distract gaze from the monitoring and treating the patient.

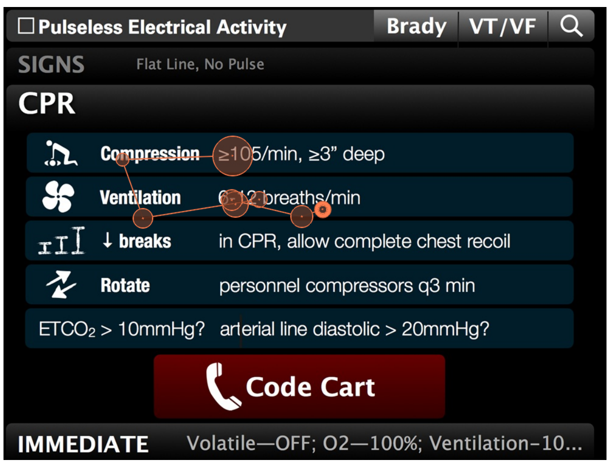


Figure 13: This aid was designed with the RapidRead principles: design patterns that yielded fast, low-variance response times, with predictable, efficient gaze paths (circles & lines)

Work on paper checklists in aviation has focused on the importance of good typography in making checklists easy and fast to read [Degani 1992]. In medicine, design advice, in the form of a checklist for making checklists [Gawande 2013], provides guidelines distilled from experience, such as using "fewer than 10 items per pause point".

The empirical work on designing crisis checklists, while expanding, is still thin. No prior work has compared alternative layout styles, and none of these guidelines have been empirically tested. This chapter has two major contributions towards these goals: a formative set of comparative measurements of information-finding tasks, and a set of design patterns found in these checklists that empirically improve performance on these tasks.

First this chapter describes RapidRead design principles, which seek to achieve those goals through structured object-action presentation and consistent design. Then, the first experiment compares five alternative checklist presentation styles from the literature: Standard Text [Ziewacz et al. 2011], Modified Standard, Color Block [Chu and Fuller 2011], Pictographic [Chu and Harrison 2012], and Dynamic Focus (see chapter 3). These aids are all focused on treating Advanced Cardiac Life Support (ACLS) medical crisis response. The within-subjects study with medical professionals (n=13) found Dynamic Focus aids to be fastest. Eye-tracking analysis showed the importance of clear visual navigation paths, anchors, and rapid scanning (see Figure 13). A second experiment compared the Dynamic Focus aids with a new design that applied the RapidRead principles to those Dynamic aids. This revision further reduced performance variation. We discuss reasons for these benefits, reflect on our combination of performance and eye-tracking data, and conclude with future work.

5.2. RapidRead Design Principles

RapidRead principles combine design patterns found in existing ACLS aids, principles derived from human perception and multitasking research, and close collaboration with medical doctors. Some of these principles were briefly sketched in

Chapter 3, and were then refined by observing other aid designs and using the analysis from the first part of the study that compares the five different procedure aid designs.

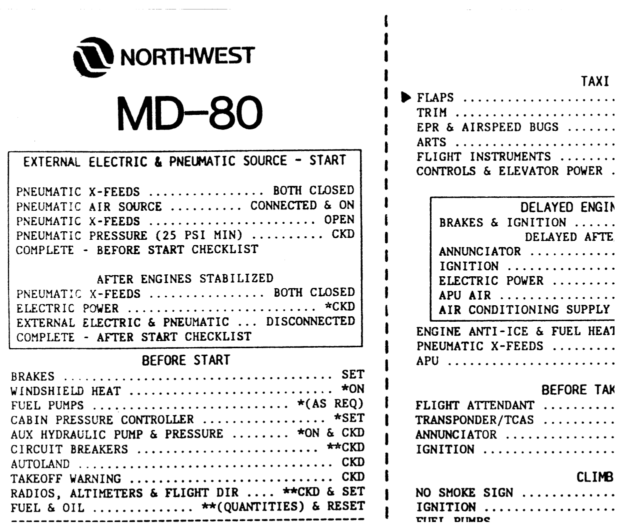
These techniques are designed to increase the speed of searching for information on demand in procedure aids. The design concept of fitting a checklist or cognitive aid step into a multi-tasking cycle we call a *step-at-a-glance* user interface.

The RapidRead design concept combines three techniques:

1. Expressing the medical information concisely in a stereotyped linguistic format called *object-action language*;
2. Mapping knowledge into graphically-defined *information patches* to increase speed of search; and
3. Dynamically adding detail near the doctor's place of focus while reducing it elsewhere, called *focus+context*.

The compact object-action language frees space on the display and visual chunks grouped for faster search by providing a short verbal handle. Visual chunks constrain visual search and this effect is amplified by the dynamic adjustment of focus+context.

Northwest Airlines MD-80 checklist



NORTHWEST

MD-80

EXTERNAL ELECTRIC & PNEUMATIC SOURCE - START

PNEUMATIC X-FEEDS .... BOTH CLOSED
PNEUMATIC AIR SOURCE .... CONNECTED & ON
PNEUMATIC X-FEEDS .... OPEN
PNEUMATIC PRESSURE (25 PSI MIN) .... CKD
COMPLETE - BEFORE START CHECKLIST

AFTER ENGINES STABILIZED
PNEUMATIC X-FEEDS .... BOTH CLOSED
ELECTRIC POWER .... *CKD
EXTERNAL ELECTRIC & PNEUMATIC ... DISCONNECTED
COMPLETE - AFTER START CHECKLIST

BEFORE START
BRAKES .... SET
WINDSHIELD HEAT .... *ON
FUEL PUMPS .... *(AS REQ)
CABIN PRESSURE CONTROLLER .... *SET
AUX HYDRAULIC PUMP & PRESSURE .... *ON & CKD
CIRCUIT BREAKERS .... **CKD
AUTOLAND .... CKD
TAKEOFF WARNING .... CKD
RADIOS, ALTIMETERS & FLIGHT DIR .... **CKD & SET
FUEL & OIL .... **(QUANTITIES) & RESET

TAXI
FLAPS ....
TRIM ....
EPR & AIRSPEED BUGS ....
ARTS ....
FLIGHT INSTRUMENTS ....
CONTROLS & ELEVATOR POWER .

DELAYED ENGIN
BRAKES & IGNITION ......
DELAYED AFTE
ANNUNCIATOR ....
IGNITION ....
ELECTRIC POWER ....
APU AIR ....
AIR CONDITIONING SUPPLY

ENGINE ANTI-ICE & FUEL HEAT
PNEUMATIC X-FEEDS ....
APU ....

BEFORE TAK
FLIGHT ATTENDANT ....
TRANSPONDER/TCAS ....
ANNUNCIATOR ....
IGNITION ....

CLIMB
NO SMOKE SIGN ....
IGNITION ....

Figure 14: Configuration checklist for a large commercial aircraft [National Transportation Safety Board, 1988]. This checklist shows an early example of object-action language.

5.2.1. Object-Action Language

This pattern is a codification of a common design pattern from aviation checklists. Figure 14 shows an example of an aviation checklist for an MD-80 airliner. It emphasizes the configuration of the airplane for take-off, landing, etc. On the left of each column is an object; on the right is an action to be taken with respect to that object, most often expressed as a configuration state to which it is to be set. For example:

BRAKES .... SET
WINDSHIELD HEAT .... ON

The language of this checklist is compact, even terse. This compactness has at least four advantages: (1) more checklist steps can be fit in a small space, (2) the checklist can be searched quickly for some setting because the objects of the search are aligned on a single page, and (3) the steps are quick to read because literary variants of the steps have been reduced to a single canonical form, and (4) the left side can be used verbally to refer to the whole in conversation.

We can use a variant of this idea in our medical checklists. For example, the more usual checklist language,

Increase FiO_2 to 100%
Verify ischemia with 12 lead EKG if possible

could be re-expressed as

FiO_2	↑100%
Ischemia	Verify
	Use 12-lead EKG.

We call this *object-action notation*.

Drug parameter sub-language. Drug dosages appear frequently in checklist statements. Misreading these statements is so consequential that it is necessary to have a special canonical way of expressing dosages. In this sublanguage, the drug name is given (even if this is a repeat of the object), followed by the dose and units (in square brackets if it is an interval), followed by special instructions like 'IV' or 'max dosage' or a complicated dosing protocol.

Calcium chloride 1g IV
Epinephrine [2~10μg/min]

Machine Parameter sub-language. Generally, the machine name will appear as the object and the action will relate to the parameter of the machine and either be of the form "Parameter = Value" or "Parameter: Action". For example:

Pacer Electrodes: Place on chest
Mode = Pacer
Current: Increase mA until capture

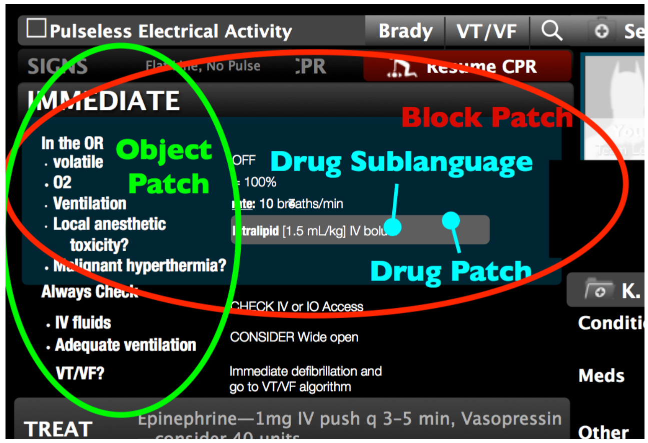


Figure 15: RapidDynamic aid principles: patches highlighted

5.2.2. Visual Information Patches

It is also important to design the aid to be rapidly searchable. Our basic concept here is to create visual patches that limit the search for target information to a smaller region. Several of these regions overlap, so different techniques are required to separate them.

Procedure blocks. Procedure blocks group a small number of procedural steps (up to five) into a block. Blocks can be of several types: *do immediately blocks*, *treatment blocks*, or *diagnostic blocks*. Procedure blocks have a subtly colored background that serves the two functions of identifying the block type and of using a low spatial frequency region to define the patch perceptually. By current theory, the human visual system has two separate pathways, a focal or "what" pathway and an ambient or “where” pathway [Ungerleider and Mishkin 1982; Tovée 2008]. The “what” pathway uses the full range of spatial frequencies whereas the ambient mode is activated by low spatial frequencies such as by stimulating large areas of the visual field [Leibowitz et al. 1984].

Drug patches. We add a gray (low spatial frequency) background under the drug parameter specification to enhance its searchability (see Figure 15).

Object patches. We wish to make a searchable patch for these, but they cross over the patches we have defined for procedure blocks. The solution is to use Tufte’s concept of *layering and separation* [Tufte 1990]; in this case, using a heavy typeface for objects together with a light typeface for other elements and separating them into two columns (see Figure 15).

5.2.3. Dynamic Focus+Context Patches

The previous techniques can be implemented in static media. The Focus+Context technique [Card 2013; Card et al. 1999; Furnas 1981] is dynamic, displaying greater detail for elements to which the doctor is attending.

5.3. Experiment 2 Part 1: Task Time Measurement

The first experiment compared five different sets of checklists in a within-subjects experiment on medical professionals. This laboratory experiment asked participants to find information embedded in procedure aids for Advanced Cardiac Life Support (ACLS) [Neumar et al. 2010]. ACLS was chosen due to its ubiquity and importance—medical doctors often are required to complete a course in ACLS before graduation, while EMTs are familiar with the related BLS (Basic Life Support).

Prior to the experiment, the design of each of the pre-existing four aid styles was analyzed. A set of design patterns, described prior to this, were formulated. Based on these design patterns we created a fifth style of aid that attempted to take the content and gross layout of the Standard Text aids [Ziewacz et al. 2011] but improve the structure using the object-action pattern and the information patch pattern.

We hypothesized that the Dynamic Focus aid would be faster than the other styles because the Dynamic Focus design pattern. reduces the amount of information displayed at one time. The other aid styles could not incorporate the Dynamic Focus design pattern because they are all constrained to paper. Furthermore, we, were unable to use this design pattern. We also hypothesized that the Updated Standard and Color Block aids would be faster than the Standard Text aids due to increased structure, and that the Pictographic aids would perform faster than the Standard Text on questions where the images were easily interpretable, but slower on other questions where the images were not easy to interpret.

5.3.1. Method

Participants

To ensure a level of understanding and familiarity with aid content (terms and abbreviations), the 13 participants comprised 2 emergency medical technicians and 11 medical doctors. For taking part in the first experiment, participants were compensated $40.

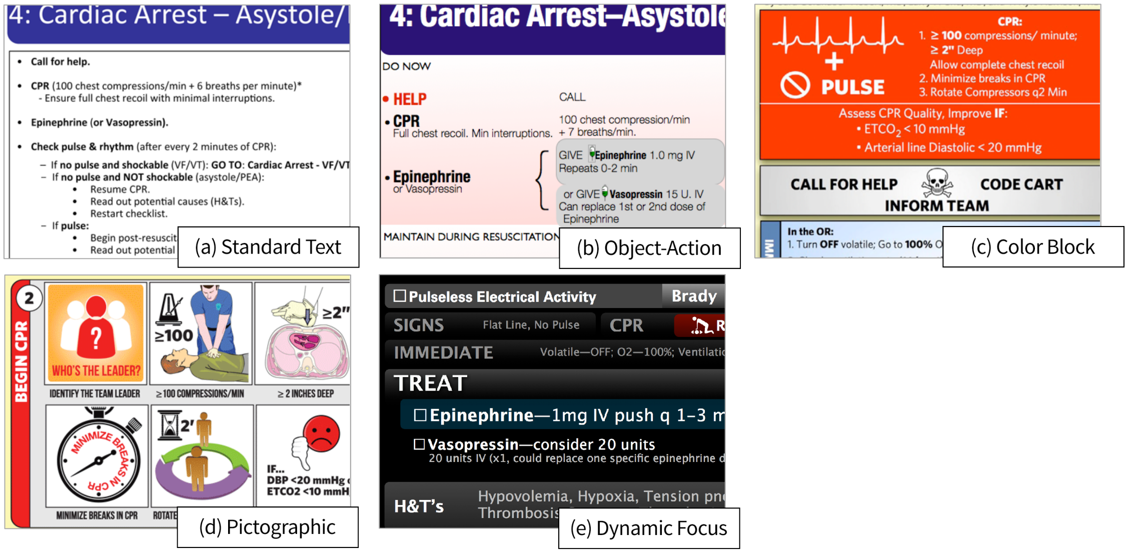


Figure 16: Asystole/Pulseless Electrical Activity aid, style comparison:
(a) ***Standard Text*****,**
(b) ***Modified Standard*****,**
(c) ***Color Block*****,**
(d) ***Pictographic*****,**
(e) ***Dynamic Focus***

Materials

The experiment compared five presentation styles:

*Standard Text***.** This set of aids was judged to be current best practice for standard paper aids [Ziewacz et al. 2011] (see Figure 16a).

Modified Standard. A variant of the standard text aids that we created to explore the effect of structured presentation. They derive most content and layout from the standard text aids, but distill their presentation into an abridged object-action format (see Figure 16b).

Color Block. These aids were developed independently of the standard text aids. The color block aids used color and visual design to delineate different conceptual chunks [Chu and Fuller 2011] (see Figure 16c).

Pictographic. These aids have similar content and wording to the color block aids, but have drastically differing visual presentation. The Pictographic aids use graphical images for each step of the checklist in addition to textual information as a way to provide visual landmarks [Chu and Harrison 2012] (see Figure 16d).

Dynamic Focus. These aids also draw their content from the Color Block aids. They streamline and structure the text, similar to the Modified Standard aids. They dynamically focus the display on the current step, showing greater detail at greater size; other steps are shown smaller and with less detail (see Figure 16e).

All styles were presented on the same 22" display. They were each given the same resolution density on screen. For example, one page of the Standard checklist aid used the same number of pixels as one page of the Modified Standard aid, and half the pixels of a two page Pictographic aid.

Procedure

Design. In a within-subjects Latin square design, participants were timed answering 15 information-lookup questions for each of five distinct styles of medical aids, totaling 75 questions.

There were five types of questions ranging from simple lookup to more difficult questions requiring some inference:

(1) Drug Parameter ("What is the correct dose for atropine?")

(2) Machine Parameter ("What rate should the pacer be set to?")
(3) Procedure Parameter ("What is the appropriate ventilation rate during CPR?")
(4) Drug Selection ("What drug and dose would you use to treat a calcium channel blocker overdose?")
(5) Procedure Diagnosis ("In pulmonary thrombosis, how do you rule out right ventricle failure?")

To ensure that responses were not remembered but found on the aid itself, question answers were altered. For example, instead of putting down the correct Epinephrine drug dosage of 1mg, we put down a similar number like 2mg. Questions spanned 4 ACLS topics: Pulseless Electrical Activity (4), Supraventricular Tachycardia (3), VT/VF (4), and Bradycardia (4). A full list of questions is available online[4].

Sequence. After a short pre-study questionnaire to record demographic information (occupation and years of experience), participants were given two example questions as a brief training. Participants were seated in a chair at a fixed distance of approximately 24" from a 22" wide-screen monitor with a resolution of 1680 pixels x 1050 pixels. Participants paced themselves using a keyboard. After reading a question, they pressed the spacebar to show the aid. Once they had found the answer, they said the answer aloud, and pressed the spacebar again to advance to the next question. The experiment measured response time for answers as the interval between spacebar presses. In addition, each session was videotaped and eye-movements were captured from the participants using a SMI RED eye-tracker. This

[4] List of questions: https://gist.github.com/icogaid/6604919

eye-tracker requires no restraint or equipment to be worn by the participant and is accurate to approximately .5~1 degree of arc.

Measures

The primary measure was the time participants took to locate the requested piece of information within the aid. Response times were compared using a fixed-effects linear model that uses participant, question, and condition. Second, we analyzed the data by comparing the fraction of responses that exceed task-relevant thresholds–10 and 20 seconds. Third, we compare variation in response times using the coefficient of variation. This metric is useful as it scales the standard deviation by the mean, allowing easy comparison between conditions. Threshold and variation analyses are important for paced tasks like crisis response and driving to measure the likelihood that an information task fits into a safe cycle time for diverting attention from the primary task [Salvucci and Taatgen 2010].

Table 2: Response times (in seconds) for different question type and aid style. The symbol ± indicates coefficient of variation, defined as the standard deviation divided by the mean.

Mean ± Coefficient of Variation	**Drug Parameter** (s)	**Machine Parameter** (s)	**Procedure Parameter** (s)	**Drug Selection** (s)	**Mean** (s)	**>10 s** (%)	**>20 s** (%)
Dynamic Focus	3.9 ± 39%	5.9 ± 52%	4.7 ± 40%	8.6 ± 59%	5.7 ± 50%	10	0.0
Color Block	6.0 ± 41%	7.3 ± 57%	8.6 ± 71%	9.4 ± 59%	8.1 ± 60%	22	4.3
Pictographic	7.7 ± 48%	8.2 ± 49%	9.3 ± 79%	9.3 ± 61%	9.0 ± 59%	30	6.8
Modified Standard	7.1 ± 58%	8.4 ± 75%	12.0 ± 85%	8.9 ± 45%	9.1 ± 70%	31	7.3
Standard Text	8.9 ± 53%	6.9 ± 56%	12.0 ± 100%	10.0 ± 59%	9.6 ± 69%	34	7.3
Mean	6.8 ± 53%	8.0 ± 61%	8.6 ± 77%	9.3 ± 55%	8.3 ± 65%	25	5.1

5.3.2. Results

The Dynamic Focus aid response times were the fastest: 41% faster (avg. 5.7s) than the Standard Text aid (avg. 9.6s). This difference was statistically significant (β

= -4.3, t(796) = -6.8, p<.001). The Color Block aid was 16% faster (β = -1.5, t(796) = -2.4, p<.05) than Standard Text. The average response times for the remaining aids were statistically indistinguishable from the Standard Text aid.

The mean time to use all aids was 8.3s, varying from 5.7s for the Dynamic Focus aid to 9.6s for the Standard Text checklist aid (see Table 2). Since long answer times are particularly dangerous, we describe what percent of trials exceed 10 or 20 seconds. The mean percent of tasks taking longer than 20 seconds was 25% and ranged from 0% for the Dynamic aid to 34% for the Standard Text checklist aid.

Answer response times were log-normally distributed: for the log-transformed distribution, skewness was 0.5 and the excess kurtosis was 0.1, both close to the values of 0 we would expect for skewness and excess kurtosis in a normal distribution, demonstrating that the distribution of the log of the data can be reasonably analyzed as a normal distribution. For all of the statistical analyses that depend on data normality, we have taken the log of the data before doing these analyses.

5.4. Discussion

Our hypothesis, that they Dynamic Focus aids would be the fastest, proved true. Our hypothesis that the Object Action aids would be faster than the Standard Text aids was not substantiated. Our hypothesis that the Color Block aids would be faster than the Standard aids was also substantiated. The Pictographic aids had mixed results in comparison to the Standard aids.

Why were the Dynamic aids so much faster?

Each of the checklist styles sought to effectively present information. Which attributes correlated with faster search? The eye traces highlight three strategies that appear to have been effective. Successful designs reduced searchers' eye movements by laying out a search path, quickly guiding them to a salient patch, or reducing the effort of digesting information once found.

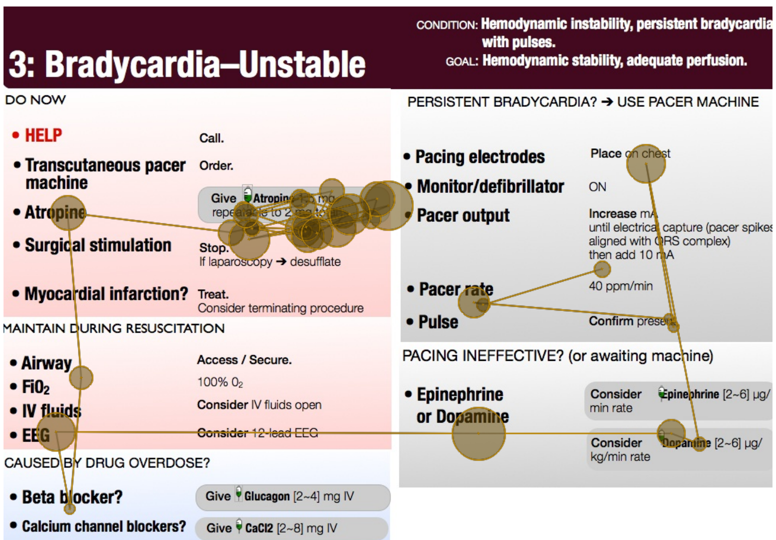


Figure 17: Structure of object column patches stands out in gaze data.

Consistent structure

Consistent with information foraging theory [Pirolli 2007], we observed that most eye traces comprised a broad scanning phase to locate the right information patch, followed by focused consumption of that patch's information. Figure 17 shows how the consistent presentation using the object-action language and object patch format speeded participants' scanning. Participants' eyes followed the object column until they found the drug name, then moved to the action column to read the dosage information. By contrast, the standard text aids have less visual structure, which required participants to exhaustively scan all of the text rather than just the anchors (see Figure 18).

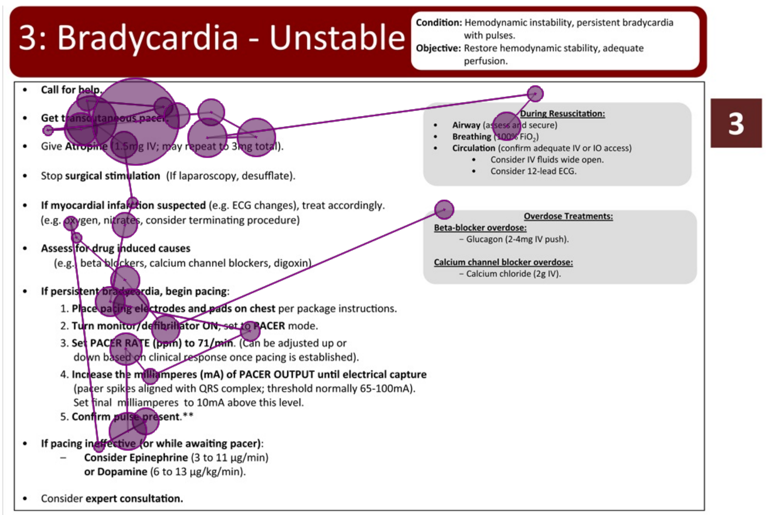


Figure 18: Low visual structure results in less gaze structure.

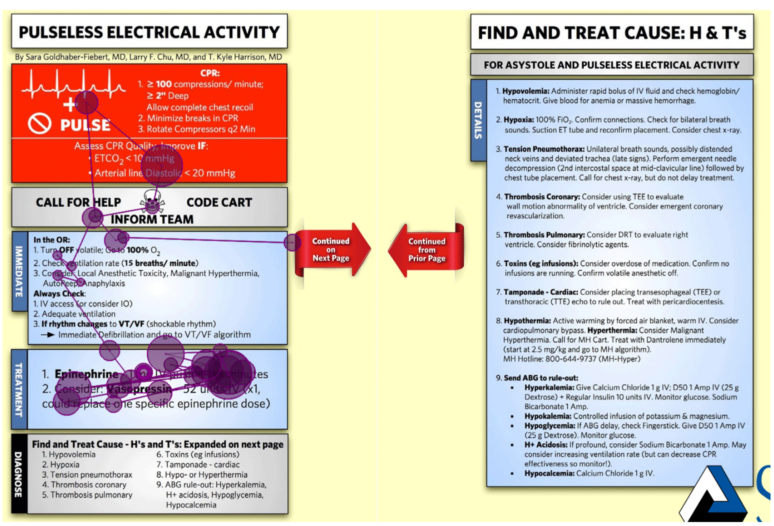


Figure 19: Gaze paths for the Color Block aid suggest that visual chunking helps guide participant'.

Clear Blocks

Information blocks help participants find information faster. By containing related information, blocks allow participants to quickly dismiss or focus on a patch. In Figure 19, the participant quickly dismissed 3 blocks before locking onto the treatment box.

Only the necessary information

Reducing the amount of information makes choices easier [Hick 1952]. In static paper layouts, there is a tradeoff between the amount of information available and the search complexity. Dynamically expanding step-relevant information and minimizing irrelevant information sped participants' search (see Figure 20).

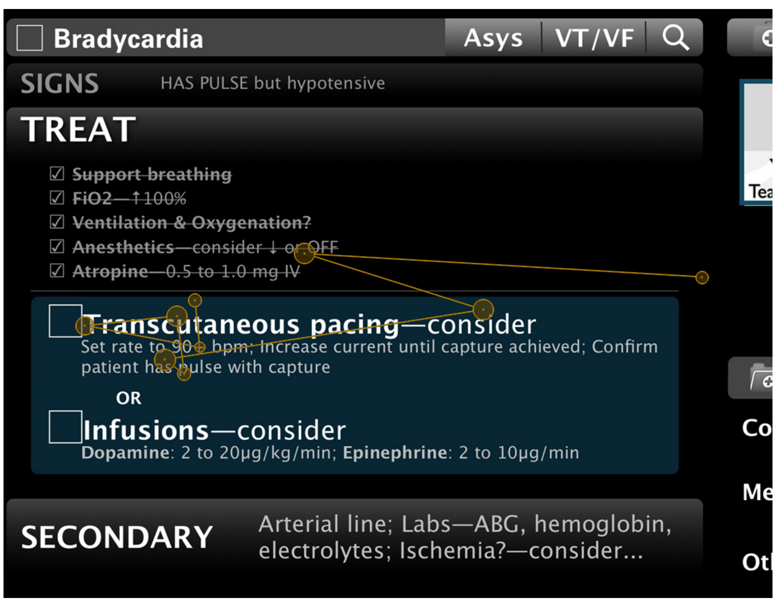


Figure 20: Dynamic Focus demonstrates fast convergence

5.4.1. Troubleshooting Cognitive Aids

This study also illuminated design flaws and opportunities for improvement in all of the aid styles. By analyzing questions with highly differential response times across the designs, we could focus on places where information design had a significant impact. In Figure 21, points along the y=x line indicate questions where response times for an aid were equivalent to the Standard Aid. This more interesting points to look at are those in the top-left or bottom-right of these graphs, because these indicate questions where a design is performs much better or much worse than the Standard aids. Here are three especially salient design issues identified with response time data and understood using the eye-tracking data. These issues highlight useful design patterns, or anti-patterns, that can be used to improve aid design.

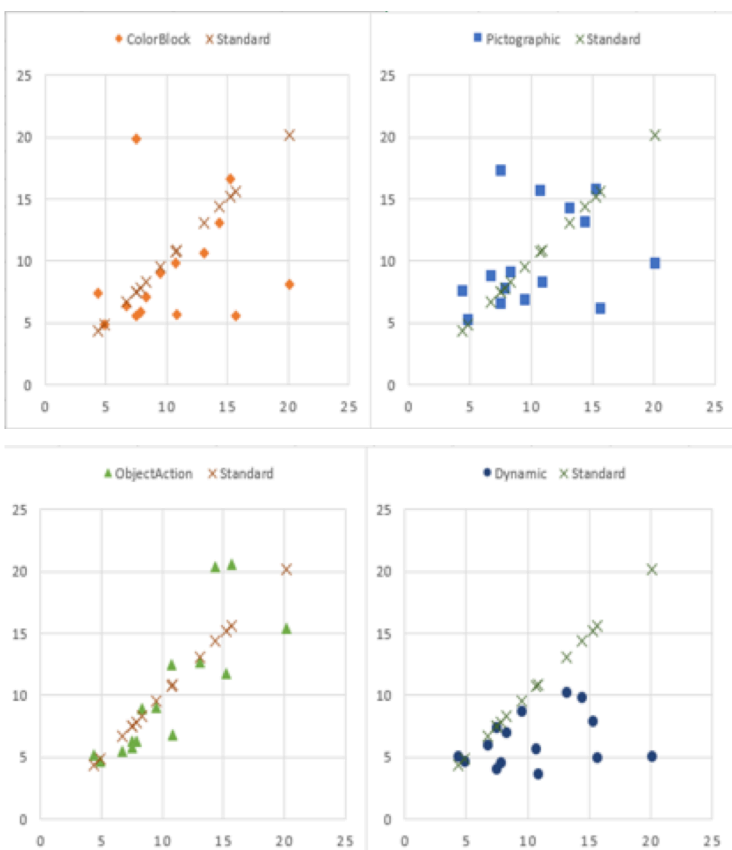


Figure 21: A comparison of mean answer times (seconds) of questions from each aid style to Standard (plotted along y=x). Points below the line y=x indicate response times faster than Standard. Dynamic (right) was fastest, but each style performed well on some questions.

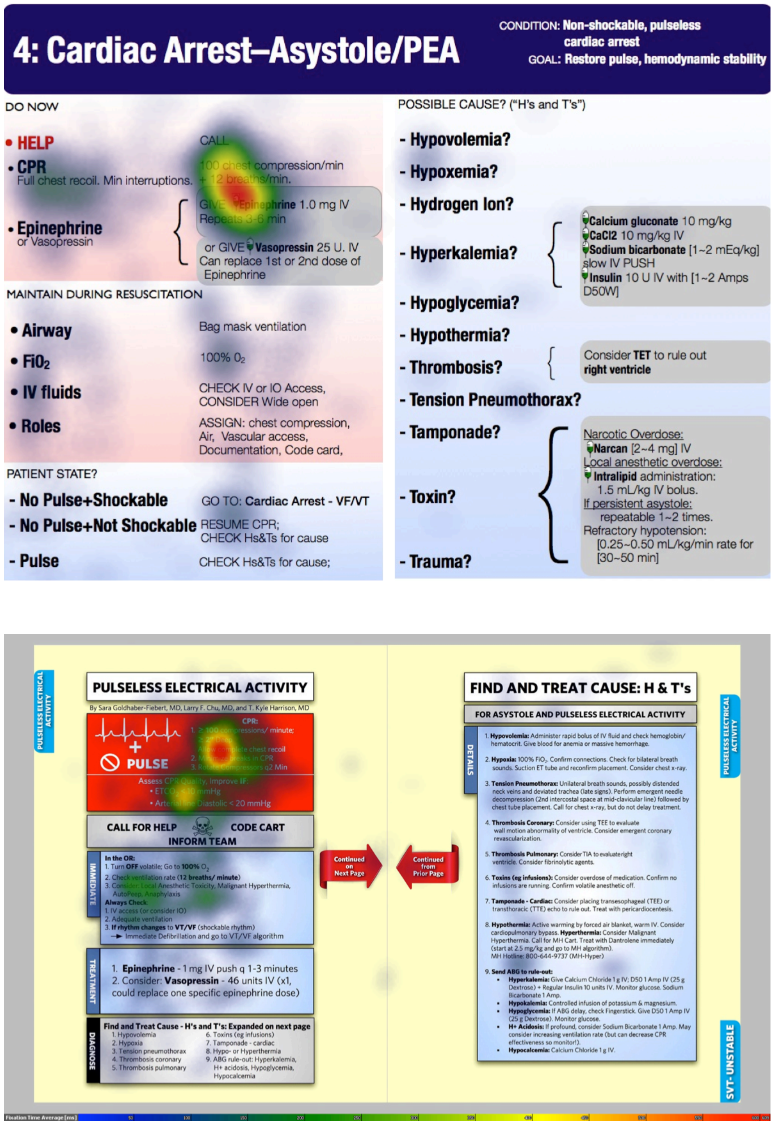


Figure 22: (top) Modified Standard: Answer aligns with eye-gaze (bottom) Color Block: Answer not aligned with primary gaze.

Finding when visual paths can help or hinder

The first design issue was with failing to group procedure information in the same patch, which resulted in participants spending long periods of time looking in the wrong place on the aid. For the question, "What is the appropriate ventilation rate during CPR for a patient in PEA?", the Modified Standard aids had average response time of 6.1s, median 5.1s, and sd of 3.0s. The Color Block had many more slow responses, with an average of 15.9s, median 11.7s, and sd 13.2s. What led to this large difference?

The heat maps reveal that the Modified Standard aid had all CPR related information in one procedure block, and that's where people spent nearly all of their time looking (see Figure 22). In contrast, participants using the Color Block aid spent most of their time looking in one procedure block which had CPR related information though the answer was in another procedure block.

Lost without an anchor

The second anti-pattern was having key information in block titles that were visually de-accentuated, which resulted in participants repeatedly scanning over the information (see Figure 23). For the question, "Patient is in unstable SVT. Should shock be synchronized or unsynchronized for a narrow complex regular rhythm?", the Dynamic aid had average response time of 8.9s, median 7.8 sec, and sd of 3.2s. The Modified Standard aid had average 18.7s, median 19.1s, and sd 6.5s.

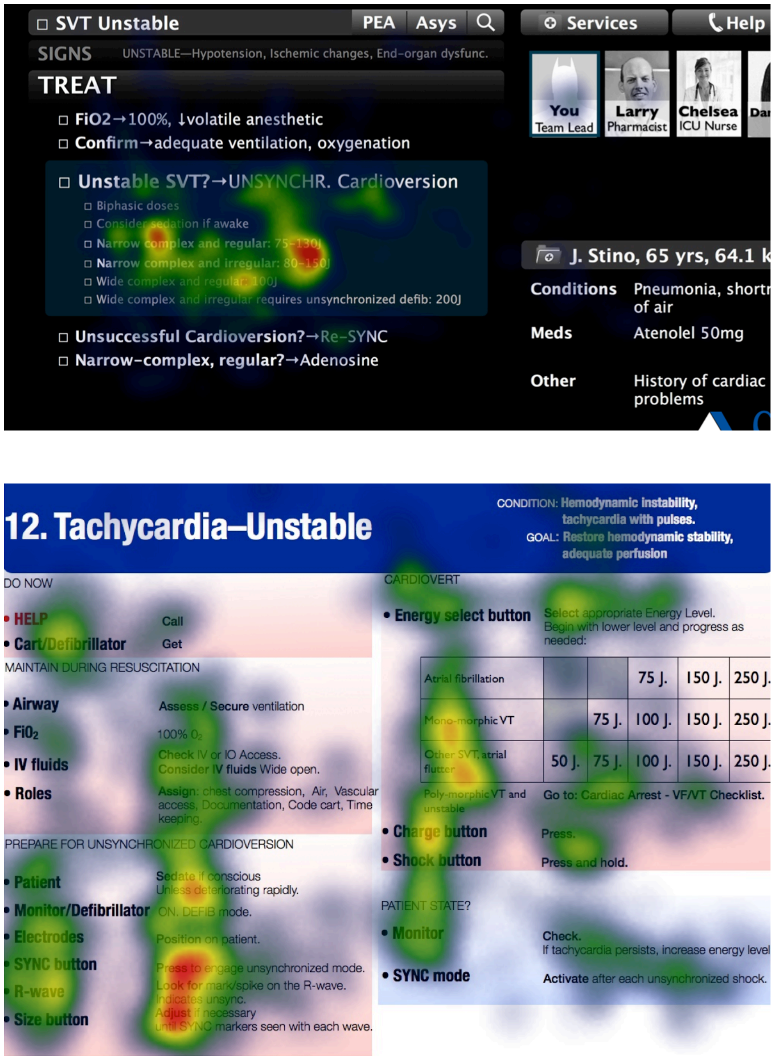


Figure 23: (top) Dynamic Focus (bottom) Modified Standard

In the Modified Standard aid, the key information—that the shock should be 'unsynchronized'—is located in a small font, all caps, and as a block title that many people missed when scanning to the larger bold items just below (see Figure 23). It also breaks one of the implicit ideas within the RapidRead, which is that all actionable information should be located in the right column of the object-action form of the procedure item and that titles and object parts of the item are used to support faster visual navigation of the aid by providing context. In the Dynamic Focus example, the line with the 'unsynch' is highlighted, has bold key terms, and the entire line is not capitalized, helping users to quickly scan and find the information. In addition, even though there are no object patches in this Dynamic aid, the key information is still phrased as an easily identifiable object and then an action. This helps users find it and understand it quickly.

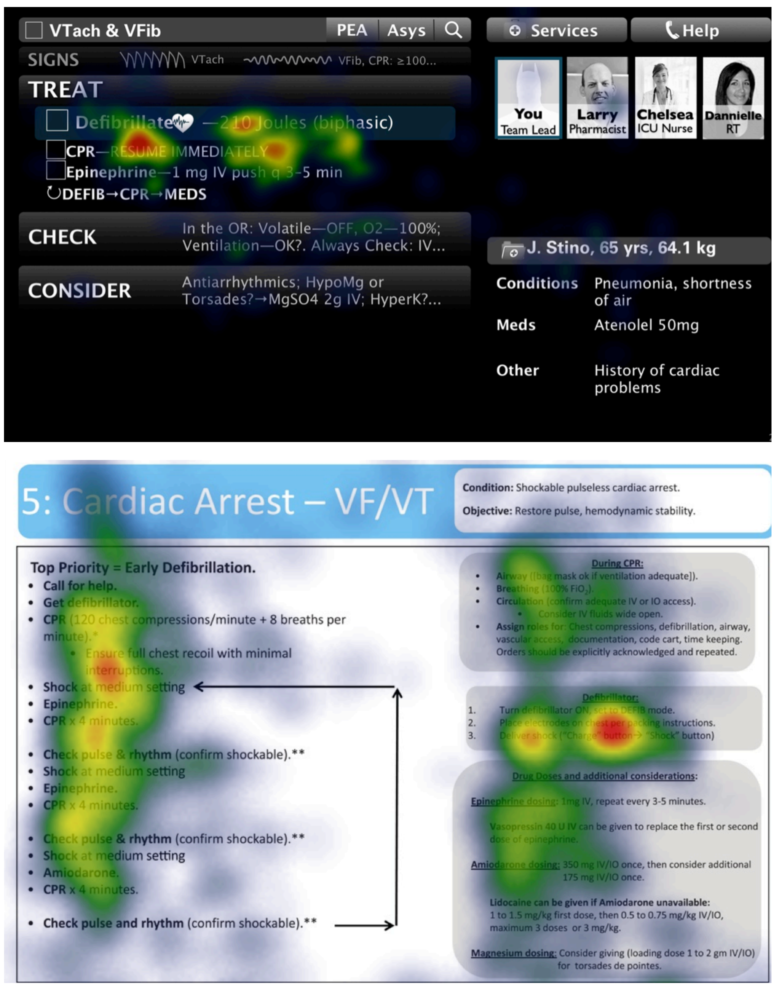


Figure 24: (top) Dynamic Focus (bottom) Standard Text

Support rapid scanning

A third design issue was the display of machine parameter settings, especially when they were repeated or split into multiple information blocks. This resulted in a new design pattern for the RapidRead principles. For the question, “How many Joules should you shock at?” the Dynamic Focus aid had average response time of 4.8s, median 4.6 sec, and sd of 1.6s. The Standard Text aid had average 18.7s, median 17.3s, and sd 13.7s.

The heat map images in Figure 24 shows a large contrast in the behavior of participants across the two conditions. When using the Dynamic aid participants efficient. They first hit the title ‘Defibrillate’ and then looked to the right to see the number of Joules. By contrast, when using the Standard Text aid, participants spent time looking in 4 separate areas. This was because machine parameter information was spread over three different sections. The largest distractor was the middle right, where there was a block of information titled ‘Defibrillator’ that didn’t contain the shock setting. The second distractor was the top right box, titled “During CPR”. The actual content was located along the left hand column. Repetition within this column seemed to hurt rather than help search time as participants would sometimes cross-check the different repetitions to make sure they were getting the right answer.

5.5. Experiment 2 Part 2: Improved Aids

The RapidRead principles were updated based on the previous three case studies and similar analysis. The machine parameter sub-language was added and the object patches pattern was validated through evaluation. This raised the question:

would re-applying these guidelines to one of the sets of aids result in improved performance?

The Dynamic Focus aid set was the best instantiation of the RapidRead principles, so we updated it to our new guidelines and created a revised version called RapidDynamic. In this version we included object patches, drug patches, the drug parameter sub-language, and the machine parameter sub-language.

5.5.1. Method

Participants

Eleven of the thirteen participants from the first experiment returned for the follow-up study. For taking part in the second experiment, participants were compensated $40.

Materials

The new RapidDynamic design was created to compare to the Dynamic. The presentation format was the same as the first experiment.

Procedure

We tested participants on a comparison between the original Dynamic design and the RapidDynamic redesign that used design principles derived from our first experiment.

5.5.2. Results

Dynamic averaged 3.7s with a sd of 4.2 and a coefficient of variance of 0.57. RapidDynamic averaged 3.1s with a sd of 0.95 and a coefficient of variance of 0.29. The difference between the means was not significant, but the difference between the variance of the two was highly significant ($F(48,48)=3.4$, $p < 0.001$).

5.5.3. Discussion

RapidDynamic was not statistically faster, but importantly had significantly reduced variance. For paced task environments like crisis medicine, this can be even more important than increasing average speed. The time difference between a task that takes 8 seconds and one that takes 10 is pretty minor, but if the task takes 30 or 60 seconds even one time it can be disastrous. By increasing the consistency of the design, our RapidRead techniques try to reduce those outliers and make aids more dependable. In the eye traces we saw the same behavioral changes between the Dynamic and RapidDynamic aids as is visible between the Standard Text and Modified Standard aids in Figure 17 and Figure 18.

Chapter 6

Dynamic Procedure Aids

6.1. Benefits of Dynamic Aids

From our studies, we note examples of ways that digital aids help.

Digital aids can track changes in best practices and protocols. Medical best practices change frequently, so even a doctor who perfectly recalls medical school knowledge may not have an up-to-date response. Take, for instance, cardiac arrest. Here is a setup and question from our study:

> *You are 10 minutes into treating a cardiac arrest. The patient's heart is in ventricular fibrillation, a heart rhythm that can be fixed by defibrillation. In the scenario your team has just shocked the patient and it looks like the patient's heart rhythm has returned to normal. What do you do next?*

Two answers are generally given here. The first is to check the patient for a pulse. This answer is given because although the patient's heart monitor shows a normal (sinus) rhythm, the electrical rhythm may not cause the heart to beat and generate a blood pressure sufficient to detect by feeling for a pulse. If the patient has a pulse, he or she does not require further treatment using this aid. If not, the patient is still in cardiac arrest. Prior to 2010, best practice was to check for pulse and rhythm changes immediately after shock, but later research showed this was not the best treatment. It is better to immediately perform 2 more minutes of post-shock CPR for *all* patients that have been in cardiac arrest, even if they have a pulse [Neumar et al. 2010]. This led to a protocol change in 2010. The current protocol answer is to always perform CPR after shocking the patient, with a possible exception when the patient has been in cardiac arrest for less than one minute.

Performing CPR before checking for a pulse (the hoped-for outcome of the shock) is both counter-intuitive and counter to previous training for many participants. The 24 participants trained in ACLS before 2010 initially learned a protocol that is no longer current. Consequently, it is likely to be performed incorrectly without a reminder. The results reflect this: 9 of the 11 participants who saw this in the Dynamic condition responded correctly; while only 3 of 10 in the paper condition and 2 of 8 in the no-aid condition responded correctly. One benefit of digital aids is that revisions can instantly propagate globally as knowledge evolves.

Digital aids can provide access to more information. Participants often forgot specifics of the protocol such as dosing, timing, joules, and appropriate ordering. A paper aid has to fit and display all of the specifics for all situations. A Dynamic aid can track the changing scenario and provide appropriate detail in real-time, without the clutter of unnecessary details.

Digital aids can reduce costs and variability of information access. Paper aids can be tough to find, easy to lose, and inconvenient to hold. Two different participants dropped the paper aids on the floor while trying to use them, multiple participants missed questions while trying to look for information in the paper aids, and some participants became so frustrated after first use that they put them down permanently.

Digital aids (and simulation) help the low performers more. An important goal of medical crisis response—and many technology scaffolds—"is to raise up the lowest performers to the level of the average performers" [Harrison 2012]. As we saw, medical students without aids performed the worst, and aids helped their performance dramatically.

Digital aids combine with simulation for effective training. This thesis introduced the Narrative Simulation approach for evaluating crisis response. Three attributes led us to this approach. First, the consistent structure of the scenario-response approach enables us to elicit situated medical responses and compare them across participants. Second, the enforced pacing maintains an element of realism in terms of timing, and helps assess and support people's performance under tight time demands. Third, Narrative Simulation is a relatively fast and cheap technique for training and evaluation. Clearly, higher-fidelity approaches also have value by helping doctors practice motor skills in a physically authentic venue. Our experience has been that simulations provide an excellent venue for introducing and evaluating digital aids. Using aids and simulation together helps both training and research. This insight builds on several decades of research into simulators for crisis response [Degani and Wiener 1993; Gaba et al. 2001], and we hope future researchers will find it valuable to build on the strategies introduced here.

Since we have discussed the training benefits of Dynamic aids, we should address a related concern: will checklists and other aids de-skill experts? People as far back as Socrates have worried that knowledge recorded on paper and other media will become a crutch that de-skills memory [Plato et al. 1961] (though it is only through recorded media do we know this view). However, with checklists as with books, this isn't a zero-sum game. Yes, people "delegate" the memory of some knowledge to recorded media when they believe they can access it later [Sparrow et al. 2011]. Given the fragile nature of memory, this is often a wise choice. Concurrently, people strengthen their information search, assessment, and integration skills—improving the quality of diagnosis and treatment.

Another worry is that checklists, whether paper or software-based, could increase error rates, or change the kinds of errors that are more likely. For example, a team leader could overfocus on a paper checklist and subsequently respond more slowly to unexpected events, or they could use the incorrect checklist. A low-ranking staff member charged with the role of reading checklists aloud [Burden et al. 2012] may feel uneasy speaking up, leading to missed steps or diagnoses. Social challenges aside, in practice checklists have shown to be useful in a number of medical tasks, even though best practices for checklist use have yet to be formalized. In crisis situations, both paper and software aids have the benefit of being non-blocking. In other words, people using aids can always choose to focus elsewhere and they will not be worse off than if they had no aids at all.

6.2. Generalizing Dynamic Aids

This thesis introduced Dynamic Procedural aids: shared displays give procedures a quickly findable location and facilitate communications and coordination for the team. Step-at-a-glance allows for rapid assimilation at minimal load of procedure steps while multi-tasking with the main task. Resources-at-a-glance allows for rapid access to resources while multi-tasking. Attention triage provides support for the allocation of attention.

The interface paradigm responds to the characteristics of complex perilous procedures, specifically operating rooms and Code Blue hospital emergencies, but its parts are abstractable and can be applied to other HCI applications. We can abstract the basic design concepts from the instantiation for this application as:

Abstraction	*Instantiation*
Shared Displays	Mirrored stadium displays using crash cart
Step-at-a-Glance	Read checklist step in a glance, simplify display, focus on current context, object/action checklist language
Resource-at-a-Glance	Access resource unit in a glance, OR team names, supply stocks, lab orders
Attention Aids	Drug timers

We could re-apply the paradigm to other suitable applications. For example, perhaps the most common paced, perilous task is driving. Using the Dynamic Aids frame to analyze a GPS display enables us to see how these same components coordinate to mitigate drivers' attentional burden. The components are as follows:

Abstraction	*Instantiation*
Shared Display	Car GPS display
Steps-at-a-Glance	Turn by turn instructions
Resource-at-a-Glance	Road names, estimated arrival time, coffee shop locations
Attention Aids	Display update, spoken turn-by-turn

GPS navigation, unlike paper maps, provides a quickly findable display that can usually be seen by both drivers and passengers. Like dpAid, input is often best delegated to the person in the support role (a nurse or passenger). Turn-by-turn navigation reveals directions with step-at-a-glance. Information readouts provide resources-at-a-glance, like estimated arrival time, current gas mileage, and potential locations to stop (for *e.g.*, gas, money, or coffee). The car's current location is displayed with a large, easily-found marker, helping to triage attention. While driving a car is not nearly as complex as performing surgery, mistakes often result in death or injury, and the role played by electronic devices in driver distraction and automobile accidents is of particular concern. GPS systems or smartphones may cause driver distraction and lead to accidents. Driving is what we might call a routine perilous procedure. There are many potential designs for reducing attentional load and other benefits. As we have done in this thesis, Narrative Simulation could be used to quickly compare such designs to find the best improvements.

6.3. The Future of Dynamic Aids

As adoption of smartphones, tablets, and heads-up displays increases, medical practice during emergency events will also continue to evolve. Smartphones can provide doctors with critical information about a patient, serve as a communication channel, and also provide cognitive aids tailored to the situation. Similarly, heads-up displays may one day replace the 20th century pager, and serve as a delivery mechanism for the private use of cognitive aids. On the other end of the visibility spectrum, wall-sized displays and pixels everywhere—from digital drapes to wearable computing—provide ways to increase shared understanding and visibility of important information. While our focus has been on medical aids, the Dynamic aid user interface paradigm was designed to be broadly useful for designing real-time assistive user-interfaces.

Deploying checklists and other cognitive aids through software has broad benefits for authoring, sharing and distributing best practices. One of the major challenges of creating excellent checklists is that someone who is an expert in both medicine and graphic design must individually craft them. This limitation prevents site-specific checklists, impedes their broader creation, and slows their revision as medical knowledge evolves. Encoding best layout practices in software enables more medical experts to participate in checklist creation and revision, and digital distribution can speed their adoption around the world. Digital aids also support automatic recording of medical procedures. Looking further into the future, Dynamic Procedure aids may help reveal new correlations between treatment and outcome. This additional information could help medical professionals make the best situation-specific decisions.

Designing tools to support crisis response can be a challenge given the paced, high-risk, multi-tasking and team-reliance of the medical domain. Digital aids offer the ability to reduce the impedance between a doctor’s needs and the information shown, to improve adoption, and to increase awareness.

Chapter 7

Conclusions and Future work

Digital checklists can add value through four key areas: being readily accessible on large screen and personal displays, being easy to rapidly assimilate, getting professional acceptance through providing additional resources, and having functionality such as timers that can help doctors with limited available attention. The RapidRead design guidelines provide specific advice on making both digital and paper checklists more consistent and faster to use.

Success is measured differently for each of these four challenges. Adoption of a successful aid will only occur when the use of that aid is both practically advantageous and culturally acceptable; therefore, aids must be both unobtrusive and efficient to use. When serving emergency responders for whom milliseconds matter, it is even more challenging to meet these practical and "cultural buy-in" goals.

Narrative Simulation provides an inexpensive way to evaluate checklists while maintaining important characteristics of the real crisis scenarios. Eye tracking provides detailed information about how the design of checklists impacts the speed and consistency with which people are able to find information.

I did this thesis work as a joint project with a second PhD student. We worked with a larger interdisciplinary team of computer scientists and doctors through a process of participatory design. Through this work we learned about the strengths of interdisciplinary work, such as combining design expertise with domain expertise, as well as the challenges, like aligning goals and expectations.

This work has implications for the future of medicine, as well as other domains where checklists can be used as a reminder or process, a safety net, or a way to scaffold expertise.

Looking forward, there remains a lot of interesting follow-up work on Dynamic Procedure aids, designing checklists, and personalizing crisis response displays.

7.1. Summary of contributions

This thesis makes contributions in three areas:

- Dynamic Procedure aids address four key issues in crisis response: ready access, rapid assimilation, professional acceptance, and limited attention. Two studies found that show how Dynamic Procedure aids perform better than paper aid styles in both Narrative Simulations and information-finding tasks.
- Based on observation work, we created the RapidRead design guidelines, which feature a consistent object-action presentation structure. A study using an eye-tracker setup compared 5 checklist layouts. It found that the presence of design patterns like object-action language, information patches, and

focus+contex improved response time and showed more structured gaze patterns.

- We developed Narrative Simulation, a scenario-driven evaluation technique. It works well for domains where usage with a system happens over a period of time, and sub-pieces of the scenario are not readily isolated without loss of context. Potential applications include testing of ubicomp or high-risk interfaces where you can't easily test in the natural setting.

7.2. Reflections on interdisciplinary work, participatory design, Narrative Simulation, rapid prototyping and doing a joint thesis

Through the process of doing research on Dynamic Procedure aids we begin to understand drawbacks and advantages of our process. We take a moment here to reflect on several of our choices.

7.2.1. Interdisciplinary work

While many interesting problems are contained within the field of computer science, computers have the potential to solve critical problems and have huge impact in many other domains. One of the critical features for making this happen is assembling research groups that contain deep knowledge in both computer science and the domain of application. In our work we have applied human computer interaction principles to the domain of medical crisis response. Our insights and successes would not have been possible without an interdisciplinary team working together. Digital checklist design for medicine requires good understanding of what computers can contribute that paper is not already doing, and it also requires good understanding of current treatment algorithms and socio-cultural best practices in medicine.

One of the reasons why our collaboration worked well is that both teams had complementary domain expertise. The doctors were experts in designing and using cognitive aids, while the computer science team had expertise in creating interactive digital applications.

In our collaboration, the computer science group was responsible for design iterations between full team design sessions, doing initial versions, bringing new perspective and problem solving techniques to the medical domain, implementing design changes, deciding on the eventual measuring stick that was used to evaluate the project, and interpreting those results. The computer science side focused a lot on how digital aids could be different from paper aids, based on a long literature in HCI of digital losing to paper when you try to replace paper. The computer science team, especially Stu Card, systematized the cognitive aids in order to simplify and standardize their creation.

Doctors contributed their domain and design expertise in a number of ways:

- Identifying and highlighting errors in observations of high-fidelity simulations. They explained both what was normal and what was abnormal in scenario response and where they had seen errors in scenario response and in cognitive aid use.
- Giving feedback on designs for digital cognitive aids. Forming opinions on how doctors might use aids based on their own experience. Deciding what pieces were critical for designs and which were distractions.

- Providing paper cognitive aids, both in terms of design and wording of those aids. These designs represented a tremendous amount of work and were the primary reason the CS team was able to quickly dive into the core issues instead of having to worry about producing the aid content from scratch. Designs also provided a comparison point for identifying design elements contributing to better aids.
- Helping design scenarios for Narrative Simulation.

Interdisciplinary work also can pose extra challenges. In particular, different research areas have different metrics that are important to them. For our medical collaborators, the important metric is very applied: in a real (or very realistic) crisis response, whether cognitive aids improve patient morbidity and reduce doctor errors. In human computer interaction, the important metric is different: how does the design of cognitive aids impact meaningful behavioral change for the doctors during crisis response? These goals are not contrary, but it can be the crux of a good collaboration to create research plans that address the concerns of both domains.

7.2.2. Participatory design with lead users

Interdisciplinary collaboration can take many forms, but ours was participatory design [Muller 2003]. Our collaborators from the medical side can be considered lead users when it comes to paper cognitive aids because they have created and used their own [Von Hippel 2006].

Our particular participatory design process was inspired by conversations with Wendy MacKay. While not explicitly based on previous work, our participatory design process closely followed the trajectory of Kristensen et al [2006].

We started with observations of medical crisis scenarios, where the computer science team was engaged with the domain. In addition, we had design meetings where the medical team engaged with the design process. Both of these types of meetings helped create “shared knowledge” between the teams as a means of supporting understanding, reflection, and design ideation [Kyng 1994].

Our design meetings took two different forms. The first form was a working session that was in a design space. In these sessions designs were presented and reflected upon, scenarios of use were discussed, and new designs were conceived of and sketched out. The second form, similar to what is called ‘Future Labs’ [Büscher et al. 2004], involved working through designs, scenarios of use, and creating new designs in the actual space where they would be used. In this case, we used the simulation lab space used for the high-fidelity simulations and observations of crisis response.

Given the nature of our interdisciplinary collaboration, participatory design was absolutely the right choice for us. Given the opportunity to repeat the experience, I would focus on improving our process in one specific way. Due to the tools involved in the design process, the computer science team did much of the between-session design work. This led to a final result that was more aligned with the problems that the computer science team was trying to solve, notably scalability and generalizability to similar domains. Von Hippel notes this challenge by observing how designs created for a broader category of users don't always optimize perfectly for individual users [Von Hippel 2006].

How could we have further put the medical team in the designers seat? Building tools to create cognitive aids earlier in the process might have accomplished this. Then many different doctors could have tried building their own dynamic and interactive aids for many different uses, leading to a series of specialized solutions that could have then informed our own designs. Another way would have been to build tools for creating evaluations. This way the doctors could have built their own Narrative Simulations and used them for their own studies, as well as helped improve the quality of our Narrative Simulations.

7.2.3. Narrative Simulation

The Narrative Simulation technique for running studies turned out to work quite well. While it would have been too costly in terms of both time and money to run all participants through high-fidelity simulations, we were able to collect data on many participants by using Narrative Simulations.

This reduction in cost comes at the price of ecological validity in two ways. First, participants were run as individuals rather than teams, so there was no way to study communication and coordination aspects of checklist use. Having the participants work as individuals rather than teams vastly simplified our analysis and gave us a good picture of standard checklist usage given. The trade-off was that we could neither understand how teamwork complicated checklist usage and created errors, nor could we look at how teamwork improved checklist usage and reduced errors.

Second, participants verbalized their treatment decisions rather than physically taking the steps required for treatment. Stated another way, we chose to test declarative knowledge rather than procedural knowledge or situated knowledge. This complicates our interpretation to some extent because what people know is dependent on their context [Hutchins 1996]. Given that doctors would, for example, make more errors remembering to restart CPR if they can't actually see the people doing CPR, we would expect them to make more errors in our study than in a real crisis situation. Our initial goal was to show that our checklists could reduce errors, so having increased errors, as a baseline, doesn't hurt our findings.

7.2.4. Two PhDs on a joint thesis

It is impossible to underestimate the amount of momentum that comes from having co-collaborator for a thesis project. The major benefit is not that work for each individual is lessened through delegation, but rather the ability to always have a sounding board and co-contributor in analysis available. This rapid iteration on ideas is far more valuable than a simple division of labor because it helps zero in on good ideas more quickly, minimizing time wasted on inferior ideas.

The main challenge in doing a joint thesis is good communication, as it is with all good collaborations, partnerships, and group projects. Here I mean good communication in two specific ways: first, in terms of the work each party has been doing independently. It is easy to see that both parties are contributing equally when the work is performed together, but when team members are working separately, balance of work can quickly come into question. One possible way to solve this is by scheduling regular check-ins to discuss progress and problems. The second way I mean communication is to make ownership explicit. I do not mean that one person works on only their own piece, although that could be one method of distributing labor, but rather that each party owns a piece of the work. This also would also mean that when a paper is being published, it is clear who is going to be the first author on that paper well prior to submission. In fact, it should be clear before the writing of the co-authored paper is started, and potentially even before the study has been run.

7.3. Implications of this work in medicine and beyond

7.3.1. Implications of these conclusions in medicine

This dissertation introduced techniques for managing complexity and improving team coordination in crisis response. As complexity and teamwork increases in medicine, we believe that the findings of this thesis will be increasingly utilized.

Paper checklists are currently a successful strategy, but in the long term it isn't a viable strategy to continue to use only paper checklists. Having books of checklists for every situation and complex machine you can encounter quickly becomes unreasonable both in terms of physical space and in terms of the time it takes to locate any one of these checklist aids. Similarly, while increasing specialization creates more manageable knowledge groupings for people to deal with, it makes it harder to spot problems borne by cross specializations and fosters increased coordination and communication challenges when treating these problems.

Computer-based cognitive aids have the ability to improve these issues. If we extrapolate from our success in ACLS, we can imagine seeing Dynamic Procedure aids in many other areas of medical practice in which we currently see checklists taking hold. However, computer-based aids are in their infancy. Understanding the design and performance of paper aids can contribute greatly to our understanding of how to design and use computer-based aids.

In addition, there is a major opportunity for Dynamic Procedure aids that can't be handled with paper aids. Dynamic Procedure aids have the unique opportunity to support cross-task cognitive aids. By using commonly linked tasks, or common exceptions to tasks, computer-based aids can seamlessly transition between the different tasks that might need to be performed. For example, you're in the middle of treating Bradycardia and your patient's heart stops. The PEA/Asystole aid is available at a touch. Or you've just arrived at a situation and you need to quickly find the Malignant Hyperthermia aid out of the hundreds of available aids. You can quickly find aids based on name or symptoms. Or, you've fixated on one diagnosis and treatment without considering another possible cause. Or, you might be treating a hypertensive patient and wish to bring up all aids matching the currently available symptoms in order to see what the next most discriminative symptom to check might be.

Checklists are the way to help, and they can help with tasks beyond mindlessly following a set of steps. Checklists are a way to augment decision making in addition to reminding doctors of all relevant steps, because sometimes the relevant step in a treatment algorithm is to decide what to do next, or to consider several treatment actions and choose the most appropriate one based on the experience and expertise of the doctor.

The end goal is to bring information to the fingertips. To make information more readily available in any kind of situation, whether it's crisis response or doing routine paperwork. The ability of "augmented" doctors to quickly find and utilize information using checklists is absolutely the future, and Dynamic Procedure aids have an opportunity to be part of the mechanism through which this happens. Perhaps more importantly, the principles that we have derived in this work are largely applicable to other information finding and checklist creation in the medical domain.

7.3.2. Where else might these conclusions be valid

While some of our conclusions are specific to the operating room context where we worked, many apply to other domains.

Rehearsing situations

Aids can be a training aid in addition to an in-situ aid. They can be a way to step through procedure steps either during downtime or en-route / just prior to a situation where you will perform the task outlined by the aid.

Checklists for training

Beyond the fact that practicing with an aid will improve your ability to use that aid in real-world situations, practicing itself may benefit from using an aid.

Cooking instructions

Cooking is simply performing a time-constrained series of tasks where you must remember amounts and preparation for each step. Dynamic Procedure aid techniques would apply very naturally to a dynamic cooking aid.

Pushing back into aviation

While a lot of insight came from looking at the work on checklists for aviation, the Dynamic checklist work synthesized and extended that work in novel ways. Using findings from the Dynamic aid work, it would be possible to go back and design improved electronic aviation aids.

Non-time constrained instructions

While Dynamic Procedure aids were designed for crisis response situations, designs would easily work for non-crisis step-by-step tasks. For example: Lego instructions, a guide to set up your own web server, a guide to using GIT to manage your code sharing, or a guide to repair your iPhone's power button.

Design for glanceble displays

One of the key design requirements for the Dynamic aid project was that the visual display was quickly glanceable, both to find information and to re-find your place. Thus, RapidRead principles should go beyond just checklists to any peripheral information displays that should be rapidly glanceable. It should be noted that there may be other principles that are useful for glanceable displays that don't fall under RapidRead principles because RapidRead was designed for a subtype of glanceable displays.

Other settings where you have one large screen display controlled for all by one person

Another aspect of Dynamic Procedure aids that can be transferred to other projects is how it uses a large-screen display for team awareness but uses an administrator as the controller. Other domains where this might be useful are control rooms (NASA, SpaceX, utility grids, power plants), software design / programming teams, or anything where team awareness is useful for maintaining good collaboration. This is in contrast with much of the related work on single display groupware where the single display is used as the main workspace for the group rather than as a peripheral information and coordination display.

7.4. Future work

The following are some open questions and some speculation as to what we expect based on our current findings.

As future work, the evaluation could be re-run in a more naturalistic setting, with a full interactive software system.

The studies presented in this dissertation measured the efficacy of novel aid styles once the right aid has been located. An important direction for future work is to measure the efficacy of techniques to locate the correct aid. We know that if the aid is showing the correct information in the correct state it is fast to find, but we haven't shown that if the aid is showing the wrong information it is no slower than current paper-based cognitive aids. Based on our trials operating the aid in the high-fidelity simulations, we believe that this is a reasonable assumption, but further evaluation of realistic interaction needs to be done. We have performed trials in high-fidelity simulations where a confederate in a reader role controlled the aid. The success of these initial studies is encouraging and requires further study.

How will Dynamic Procedure aids compare to existing paper designs in high-fidelity simulations?

We believe that the Dynamic Procedure aids will perform better than paper aids in high-fidelity simulations. Paper aids improve performance, and there is a correlation between the amount paper aids are used and the amount they help. Since we've designed and evaluated how fast and easy it is to use the Dynamic Procedure aids, we believe that this will translate to the full high-fidelity simulations. One of the major hurdles will be to make sure that the aids are on point, and easy to keep that way.

One larger goal would be the construction of a predictive theory capable of estimating time required based on a model of user behavior and visual design. This would be especially useful for rapidly supporting diverse devices.

While we have developed some theory around how design impacts performance in crisis domains, it would be extremely useful to develop further theory on how we can model user behavior in relation to visual design characteristics as a way to estimate time to find, understand, and act on information within cognitive aids. This would allow automated testing of cognitive aids and even optimization of cognitive aid designs. This builds on work on predicting eye-movements [Foulsham and Underwood 2008] and predicting interaction behavior performance [Card and Moran 1986].

Will personalization of digital aids help adoption or performance?

We have discussed literature which indicates that performance suffers when checklist are implemented without respect for the culture in which they will be used

[Verdaasdonk et al. 2009]. We think that straightforward modifications to the procedures in the Dynamic Procedure aids to match local or personal practice will make it much easier to transition the culture and workflow of the hospital into using the aids. Matching best practice with all those personal lists, or keeping those personal lists up to date when best practice changes, will require some amount of local overhead. But automatically changing lists that haven't been personalized, and requiring integrations of new information into lists that have been personalized can minimize this cost.

How does performance in Narrative Simulations relate to real crisis performance?

One primary concern with Narrative Simulations evaluation is the risk that differences the simulation and crisis response will change how aids are used. While we believe the Narrative Simulations capture the important characteristics relevant to the doctor using the cognitive aid, future work should test this explicitly.

A similar line of work has been done for high-fidelity simulations, which found that stress levels are similar between real crisis situations and their simulation counter-parts [Kharasch et al. 2011; Bong et al. 2010]. This provides one metric to compare our Narrative Simulations to high-fidelity simulations. Another useful comparison would be to compare individual performance between them and see if there is a correlation.

While not a direct measure of comparison between types of simulations, it would also be interesting to see if repeated exposure to Narrative Simulations improved performance during high-fidelity simulations.

Dynamic Procedure aids are fast to use when they are showing the correct information, but is it difficult to keep them always showing the right information?

While qualitative evaluations were done in high-fidelity simulations, our Narrative Simulations and information-finding tasks did not directly measure the interaction cost of our Dynamic design. Specifically, we did not look at the cost of finding information when the Dynamic aid was not showing the information that the participant was looking for. This is also a concern when thinking about page turning cost for paper designs that span multiple pages in order to display more information or display the information less densely.

Informally, in the high-fidelity simulations it was straight-forward to keep the right information displayed when it was needed, as was assumed in the studies. Future work should exercise the breadth of aid navigation more thoroughly. Another useful measure would be to repeat the Narrative Simulation studies where we asked participants to control the interface directly rather than having it implicitly controlled by a teammate operator. While this wouldn't match the real world as closely, it would give a reasonable evaluation of how difficult it is to control the system and keep the display showing the correct information.

How might personal heads-up displays help coordinate teams using Dynamic Procedure aids?

There are two continuing challenges with our current design that may be solved by utilizing personal heads-up displays. First, the large screen needs to show all information relevant to all team members, or barring that, the most important information across a set of team members that are using the aid. However, personally worn heads-up displays, such as Google Glass, provide an opportunity to augment the large screen display with a small amount of personalized information. For the team leader, this might be the next step from the checklist, for the pharmacist this might be the next drug and dosage that needs to be drawn up, and for the person doing CPR the heads-up display might show a measure of their CPR quality.

The second opportunity is for notifications. It is very challenging for a large screen display designed as a peripheral support tool to the main task to capture attention. The system must rely on the doctors checking in regularly to see things like the next step or the drug timers that keep track of when dosages are due. With wearable computers, such as heads-up displays, you can always be in the doctor's field of vision, or have an audio output that only one person can hear. This means that you can have more assurance that your alerts will be seen while they are still relevant without disrupting the entire team.

References

ACLS-ALGORITHMS, 2012. MegaCode.

ADAMCZYK, P.D. AND BAILEY, B.P., 2004. If Not Now, When?: The Effects of Interruption at Different Moments Within Task Execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, New York, USA: ACM Request Permissions, pp. 271–278.

AGRAWALA, M. ET AL., 2003. Designing effective step-by-step assembly instructions. *ACM SIGGRAPH 2003 Papers on - SIGGRAPH '03*, p.828.

ARONSKY, D., JONES, I., LANAGHAN, K. AND SLOVIS, C., 2008. Supporting patient care in the emergency department with a computerized whiteboard system. *Journal of the American Medical Informatics Association*, pp.184–195.

ARRIAGA, A.F. ET AL., 2013. Simulation-based trial of surgical-crisis checklists. *The New England journal of medicine*, 368(3), pp.246–53.

BAAYEN, R.H., DAVIDSON, D.J. AND BATES, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), pp.390–412.

BACHOUR, K. AND KAPLAN, F., 2009. An Interactive Table for Supporting Participation Balance in Face-to-Face Collaboration. , pp.1–10.

BAILEY, B.P. AND IQBAL, S.T., 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *Transactions on Computer-Human Interaction (TOCHI)*, 14(4), pp.1–28.

BAUDISCH, P. ET AL., 2006. Phosphor: explaining transitions in the user interface using afterglow effects. In *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*. New York, New York, USA: ACM Press, pp. 169–178.

BEDERSEN, B.B., 2000. Fisheye Menus. In M. S. Ackerman & K. Edwards, eds. *Proceedings of the 13th annual ACM Symposium on User Interface Software and Technology*. San Diego, California, pp. 217–225.

BERGEN, B., MEDEIROS-WARD, N., WHEELER, K., DREWS, F. AND STRAYER, D., 2013. The crosstalk hypothesis: why language interferes with driving. *Journal of Experimental Psychology: General*, 142(1), pp.119–30.

BIRNHOLTZ, J., RANJAN, A. AND BALAKRISHNAN, R., 2010. Providing Dynamic Visual Information for Collaborative Tasks: Experiments With Automatic Camera Control. *Human-Computer Interaction*, 25(3), pp.261–287.

BONG, C.L., LIGHTDALE, J.R., FREDETTE, M.E. AND WEINSTOCK, P., 2010. Effects of simulation versus traditional tutorial-based training on physiologic stress levels among clinicians: a pilot study. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 5(5), pp.272–8.

BOORMAN, D., 2001. Safety benefits of electronic checklists: An analysis of commercial transport accidents. In *Proceedings 11th Interantion Symposium on Aviation Psychology*. Columbus, USA, pp. 5–8.

BRANDT, J., DONTCHEVA, M., WESKAMP, M. AND KLEMMER, S.R., 2010. Example-centric programming: integrating web search into the development environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

BRENNAN, T.A. ET AL., 1991. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *The New England Journal of Medicine*, 13(2), pp.370–376.

BROWN, J.S. AND WEISER, M., 1996. Designing Calm Technology. *PowerGrid Journal*.

BRUMBY, D.P., HOWES, A. AND SALVUCCI, D.D., 2007. A cognitive constraint model of dual-task trade-offs in a highly dynamic driving task. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp.233–242.

BURDEN, A.R., CARR, Z.J., STAMAN, G.W., LITTMAN, J.J. AND TORJMAN, M.C., 2012. Does every code need a "reader?" improvement of rare event management with a cognitive aid "reader" during a simulated emergency: a pilot study. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 7(1), pp.1–9.

BURIAN, B.K., 2006. Design Guidance for Emergency and Abnormal Checklists in Aviation. In *Proceedings of the Human Factors and Ergonomics Society*. San Francisco, CA, pp. 1–6.

BURIAN, B.K., 2004. Emergency and abnormal checklist design factors influencing flight crew response: A case study. In *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics*.

BURIAN, B.K., BARSHI, I. AND DISMUKES, K., 2005. *The Challenge of Aviation Emergency and Abnormal Situations*, Moffett Field, California.

BURIAN, B.K.B.K. AND D, P., 2006. Aeronautical emergency and abnormal checklists: expectations and realities. In *Human Factors*. San Francisco, USA: Human Factors and Ergonomics Society, pp. 101–105.

BURTSCHER, M.J. ET AL., 2011. Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. *British Journal of Anaesthesia*, 106(6), pp.801–806.

BÜSCHER, M., ERIKSEN, M.A., KRISTENSEN, J.F. AND MOGENSEN, P.H., 2004. Ways of grounding imagination. *Proceedings of the eighth conference on Participatory design Artful integration interweaving media materials and practices PDC 04*, Toronto, O, p.193.

CANNON-BOWERS, J.A., SALAS, E. AND CONVERSE, S., 1993. Shared mental models in expert team decision making. In *Individual and Group Decision Making: Current Issues*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, pp. 221–246.

CARD, S.K., 2013. Commentary on: Spence, Robert and Apperley, Mark (2013): Bifocal Display. In M. Soegaard & R. F. Dam, eds. *The Encyclopedia of Human-Computer Interaction*. The Interaction Design Foundation.

CARD, S.K., MACKINLAY, J.D. AND SHNEIDERMAN, B., 1999. *Readings in information visualization: using vision to think*, San Francisco: Morgan Kaufmann Publishers Inc.

CARD, S.K. AND MORAN, T.P., 1986. User technology—from pointing to pondering. In *Proceedings of the ACM Conference on The History of Personal Workstations*. pp. 183–198.

CARD, S.K., MORAN, T.P. AND NEWELL, A., 1983. *The Psychology of Human-Computer Interaction*, Taylor & Francis.

CENTER FOR DISEASE CONTROL, 2005. *ICD-9-CM International Classification of Diseases* 9th ed., Wolrd Health Organization.

CHISHOLM, C.D., COLLISON, E.K., NELSON, D.R. AND CORDELL, W.H., 2000. Emergency Department Workplace Interruptions: Are Emergency Physicians

"'Interrupt-driven'" and "'Multitasking'"? *Academic Emergency Medicine*, 7(11), pp.1239–1243.

CHU, L. AND FULLER, A., 2011. *Appendix C: Crisis Management Cognitive Aids in Manual of Clinical Anesthesia*, Lippincott.

CHU, L. AND HARRISON, T.K., 2012. *ACLS Emergency Aids*,

CLARK, H.H., 1996. *Using language*, Cambridge: Cambridge University Press.

CLARK, H.H. AND BRENNAN, S.E., 1991. Grounding in Communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley, eds. *Perspecitves on Socially Shared Cognition*. Washington, D.C.: American Psychologcal Association.

CUTRELL, E., CZERWINSKI, M. AND HORVITZ, E., 2000. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance. *Proceedings of INTERACT*, pp.1–7.

DAVIS, P., LAY-YEE, R., BRIANT, R., ALI, W., SCOTT, A. AND SCHUG, S., 2002. Adverse events in New Zealand public hostpitaals I: Occurrence and impact. *The New Zealand Medical Journal*, 115(1167), p.U271.

DEGANI, A., 1992. *On the typography of flight-deck documentation*, Moffett Field, California: NASA-Ames Research Center.

DEGANI, A. AND WIENER, E.L., 1993. Cockpit Checklists: Concepts, Design, and Use. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(2), pp.345–359.

DEGANI, A. AND WIENER, E.L., 1990. *Human Factors of Flight-Deck Checklists : The Normal Checklist*, Ames Research Center.

DEKKER, S., 2011. *Patient Safety: A Human Factors Approach*, Boca Raton, FL: CRC Press.

DISMUKES, K. AND NOWINSKI, J., 2007. Prospective Memory , Concurrent Task Management , and Pilot Error. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik, eds. *Attention: From Theory to Practice*. Oxford: Oxford University Press, pp. 225–236.

DONCHIN, Y. ET AL., 1995. A look into th nature and causes of human errors in the ientensive care unit. *Critical Care Medicine*, 23(2), pp.294–300.

DYM, C.L., LITTLE, P. AND ORWIN, E., 2013. *Engineering Design: A Project-Based Introduction*, Wiley.

ENTIN, E.E. AND SERFATY, D., 1999. Adaptive Team Coordination. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 41(2), pp.312–325.

ERICSSON, K. A AND LEHMANN, A C., 1996. Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of psychology*, 47, pp.273–305.

FLIN, R., PATEY, R., GLAVIN, R. AND MARAN, N., 2010. Anaesthetists' non-technical skills. *British Journal of Anaesthesia*, 105(1), pp.38–44.

FOULSHAM, T. AND UNDERWOOD, G., 2008. What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 8, pp.1–17.

FOURCADE, A., BLACHE, J.-L., GRENIER, C., BOURGAIN, J.-L. AND MINVIELLE, E., 2012. Barriers to staff adoption of a surgical safety checklist. *BMJ Quality & Safety*, 21(3), pp.191–7.

FRISCH, S., FÖRSTL, S., LEGLER, A., SCHÖPE, S. AND GOEBEL, H., 2012. The interleaving of actions in everyday life multitasking demands. *Journal of Neuropsychology*, 6(2), pp.257–69.

FURNAS, G.W., 1981. *The FISHEYE view : a new look at structured files*, Murray Hil, NJ, USA: Bell Laboratories.

GABA, D.M., 2011a. Have we gone too far in translating ideas from aviation to patient safety? No. *BMJ*, 342(jan11 4), pp.c7310–c7310.

GABA, D.M., 2007. Out of this nettle, danger, we pluck this flower, safety: healthcare vs. aviation and other high-hazard industries. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 2(4), pp.213–7.

GABA, D.M., 2011b. Training and nontechnical skills: the politics of terminology. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 6(1), pp.8–10.

GABA, D.M., FISH, K.J. AND HOWARD, S.K., 1994. *Crisis Management in Anesthesiology*, New York: Churchill Livingstone.

GABA, D.M., HOWARD, ST.K., FISH, K.J., SMITH, B.E. AND SOWB, Y. A., 2001. Simulation-Based Training in Anesthesia Crisis Resource Management (ACRM): A Decade of Experience. *Simulation & Gaming*, 32(2), pp.175–193.

GAWANDE, A., 2013. Project Check - http://www.projectcheck.org/checklist-for-checklists.html.

GAWANDE, A., 2009. *The Checklist Manifesto: How to Get Things Right*, New York: Metropolitan Books.

GERGLE, D. AND CLARK, A.T., 2011. See what i'm saying?: using Dyadic Mobile Eye tracking to study collaborative reference. *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp.435–444.

GREEN, P., 1999. The 15-second rule for driver information systems. In *Proceedings of the ITS America Ninth Annual Meeting*.

GREEN, P., 2002. Traffic Safety. In R. E. Dewar & P. L. Olson, eds. *Human Factors in Traffic Safety*. Tucson, AZ, USA: Lawyers & Judges Publishing Co.

HALES, B.M. AND PRONOVOST, P.J., 2006. The checklist--a tool for error management and performance improvement. *Journal of Critical Care*, 21(3), pp.231–5.

HARRISON, T.K., 2012. personal communication.

HARRISON, T.K., MANSER, T., HOWARD, S.K. AND GABA, D.M., 2006. Use of Cognitive Aids in a Simulated Anesthetic Crisis. *Anesthesia & Analgesia*, 103(3), pp.551–556.

HAYNES, A.B. ET AL., 2009. A surgical safety checklist to reduce morbidity and mortality in a global population. *The New England journal of medicine*, 360(5), pp.491–9.

HEALEY, A.N., SEVDALIS, N. AND VINCENT, C.A., 2006. Measuring intra-operative interference from distraction and interruption observed in the operating theatre. *Ergonomics*, 49(5-6), pp.589–604.

HEATH, C., SVENSSON, M.S., HINDMARSH, J., LUFF, P. AND VOM LEHN, D., 2002. Configuring Awareness. *Computer Supported Cooperative Work (CSCW)*, 11(3-4), pp.317–347.

HICK, W.E., 1952. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1), pp.11–26.

VON HIPPEL, E., 2006. *Democratizing Innovation*, The MIT Press.

HORN, R.E., 1990. Mapping Hypertext: the analysis, organization, and display of knowledge for the next generation of on-line text and graphics. *Choice Reviews Online*, 28(02), pp.28–999.

HORREY, W. AND WICKENS, C., 2007. In-vehicle glance duration: Distributions, tails and a model of crash risk. *Transportation Research Record: …*, pp.1–16.

HORVITZ, E., KADIE, C. AND PAEK, T., 2003. Models of attention in computing and communication. *COMMUNICATIONS OF THE ACM*.

HUNZIKER, S. ET AL., 2011. Teamwork and leadership in cardiopulmonary resuscitation. *Journal of the American College of Cardiology*, 57(24), pp.2381–8.

HUTCHINS, E., 1996. *Cognition in the Wild (Bradford Books)*, A Bradford Book.

HUTCHINS, E., 1995. How a cockpit remembers its speeds. *Cognitive Science*, 19(3), pp.265–288.

IQBAL, S.T. AND BAILEY, B.P., 2005. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. *CHI EA '05: CHI '05 Extended Abstracts on Human Factors in Computing Systems*.

IQBAL, S.T. AND BAILEY, B.P., 2010. Oasis. *ACM Transactions on Computer-Human Interaction*, 17(4), pp.1–28.

JAFARINAIMI, N., FORLIZZI, J., HURST, A. AND ZIMMERMAN, J., 2005. Breakaway: an ambient display designed to change human behavior. *CHI EA '05: CHI '05 Extended Abstracts on Human Factors in Computing Systems*.

JAMES, J.T., 2013. A new, evidence-based estimate of patient harms associated with hospital care. *Journal of Patient Safety*, 9(3), pp.122–8.

JEUNG, H., CHANDLER, P. AND SWELLER, J., 1997. The Role of Visual Indicators in Dual Sensory Mode Instruction. *Educational Psychology*, 17(3), pp.329–345.

KAHNEMAN, D., 2011. *Thinking Fast and Slow*, Farrar, Straus, and Giroux.

KALNIKAITÉ, V. AND WHITTAKER, S., 2007. Software or wetware?: discovering when and why people use digital prosthetic memory. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*.

KENDELL, J. AND BARTHRAM, C., 1998. Revised checklist for anaesthetic machines. *Anaesthesia*, 53(9), pp.887–90.

KHARASCH, M., AITCHISON, P., PETTINEO, C., PETTINEO, L. AND WANG, E.E., 2011. Physiological stress responses of emergency medicine residents during an immersive medical simulation scenario. *Disease-a-month : DM*, 57(11), pp.700–5.

KOHN, L.T., CORRIGAN, J. AND DONALDSON, M.S., 2000. *To Err is Human: Building a Safer Health System*, Washington, D.C.

KOLBE, M., BURTSCHER, M., MANSER, T., KÜNZLE, B. AND GROTE, G., 2011. The Role of Coordination in Preventing Harm in Healthcare Groups: Research Examples from Anaesthesia and an Integrated Model of Coordination for Action Teams in Health Care. In M. Boos, ed. *Coordination in Human and Primate Groups*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 75–92.

KRISTENSEN, M., KYNG, M. AND PALEN, L., 2006. Participatory design in emergency medical service: designing for future practice. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p.161.

KYNG, M., 1994. Scandinavian design: users in product development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 3–9.

LAYTON, P., 2003. *Towards Managing Uncertainty: Coupling Experimentation with Rapid Prototyping*, Aerospace Centre.

LEAPE, L.L. ET AL., 1991. The nature of adverse events in hospitalized patients: Results of the Harvard Medical Practice Study II. *The New England Journal of Medicine*, 324(6), pp.377–384.

LEIBOWITZ, H.W., SHUPERT, C.W. AND POST, R.B., 1984. The two modes of visual processing: Implications for spatial orientation. In *Peripheral vision horixontal display (PVHD), proceedings of a conference held at NASA Ames research center*. Edwards, California, pp. 41–44.

MAGLIO, P.P. AND CAMPBELL, C.S., 2000. Tradeoffs in displaying peripheral information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Request Permissions.

MAKARY, M.A. ET AL., 2006. Operating room briefings: working on the same page. *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources*, 32(6), pp.351–5.

MANKINS, J., 1995. Technology readiness levels. *White Paper, April*, pp.4–8.

MANSER, T., HARRISON, T.K., GABA, D.M. AND HOWARD, S.K., 2009. Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesthesia & Analgesia*, 108(5), pp.1606–15.

MATHIEU, J.E., HEFFNER, T.S., GOODWIN, G.F., SALAS, E. AND CANNON-BOWERS, J.A., 2000. The influence of shared mental models on team process and performance. *The Journal of Applied Psychology*, 85(2), pp.273–83.

MATTHEWS, T., HSIEH, G. AND MANKOFF, J., 2009. Evaluating Peripheral Displays. *Human-Computer Interaction Series*, (Chapter 19), pp.447–472.

McCONNELL, D.J., FARGEN, K.M. AND MOCCO, J., 2012. Surgical checklists: A detailed review of their emergence, development, and relevance to neurosurgical practice. *Surgical Neurology International*, 3(1), p.2.

MENTIS, H.M., O'HARA, K., SELLEN, A. AND TRIVEDI, R., 2012. Interaction proxemics and image use in neurosurgery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, New York, USA: ACM Press, p. 927.

MONK, C. A, TRAFTON, J.G. AND BOEHM-DAVIS, D. A, 2008. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology*, 14(4), pp.299–313.

MONSELL, S., 2003. Task switching. *Trends in Cognitive Sciences*, 7(3), pp.134–140.

MULLER, M., 2003. Participatory design: the third space in HCI. *Human-Computer Interaction: Development Process*, 4235, pp.1–70.

NEALE, G., WOLOSHYNOWYCH, M. AND VINCENT, C., 2001. Exploring the causes of adverse events in NHS hospital practice. *Journal of the Royal Society of Medicine*, 94(7), pp.322–330.

NESTEL, D., WALKER, K., SIMON, R., AGGARWAL, R. AND ANDREATTA, P., 2011. Nontechnical skills: an inaccurate and unhelpful descriptor? *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 6(1), pp.2–3.

NEUMAR, R.W. ET AL., 2010. Part 8: adult advanced cardiovascular life support: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*, 122(18 suppl 3), pp.S729–S767.

NYSSEN, A.-S., LARBUISSON, R., JANSSENS, M., PENDEVILLE, P. AND MAYNÉ, A., 2002. A comparison of the training value of two types of anesthesia simulators: computer screen-based and mannequin-based simulators. *Anesthesia & Analgesia*, 94(6), pp.1560–5, table of contents.

PATTERSON, E.S., 2007. Communication Strategies From High-reliability: Translation is Hard Work. *Annals of Surgery*, 245(2), pp.170–172.

PATTERSON, E.S., RENDER, M.L. AND EBRIGHT, P.R., 2002. Repeating human performance themes in five health care adverse events. In *Human Factors And Ergonomics*. SAGE Publications.

PHILLIPS, D.P. AND BARKER, G.E.C., 2010. A July spike in fatal medication errors: a possible effect of new medical residents. *Journal of general internal medicine*, 25(8), pp.774–9.

PIROLLI, P., 2007. *Information Foraging Theory: Adaptive Interaction with Information*, Oxford University Press.

PLATO, HAMILTON, E. AND CAIRNS, H., 1961. *The collected dialogues of Plato, including the letters*, Princeton, NJ, USA: Princeton University Press.

PRONOVOST, P. ET AL., 2006. An intervention to decrease catheter-related bloodstream infections in the ICU. *The New England journal of medicine*, 355(26), pp.2725–32.

DE REE, H., 1991. The emergency checklist: testing various layouts. In *Proceedings of the International Symposium on Aviation Psychology*. Columbus, OH, USA: The Ohio State University, pp. 160–164.

REHMANN, J.T., STEIN, E.S. AND ROSENBERG, B.L., 1983. Subjective Pilot Workload Assessment. *Human Factors*, pp.297–307.

ROCHLIN, G.I., LA PORTE, T.R., ROBERTS, K.H. AND LAPORTE, T.R., 2005. The Self-Designing High-Reliability Organization : Aircraft Carrier Flight Operations at Sea. *Naval War College Review*, 40(4), pp.75–90.

SALVUCCI, D.D. AND TAATGEN, N.A., 2010. *The Multitasking Mind (Cognitive Models and Architectures)*, Oxford University Press, USA.

SALVUCCI, D.D. AND TAATGEN, N.A., 2008. Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115(1), pp.101–130.

SANBONMATSU, D.M., STRAYER, D.L., MEDEIROS-WARD, N. AND WATSON, J.M., 2013. Who multi-tasks and why? Multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *PloS ONE*, 8(1), p.e54402.

SANDERS, A., 1984. Ten symposia on attention and performance: Some issues and trends. *Attention and performance X: Control of language processes*, pp.3–13.

SARCEVIC, A., MARSIC, I. AND BURD, R.S., 2010. Does size and location of the vital signs monitor matter? A study of two trauma centers. In *Proceedings of the American Medical Informatics Association 2010 Annual Symposium (AMIA 2010)*. pp. 707–11.

SEAGULL, F.J., WICKENS, C.D. AND LOEB, R.G., 2001. WHEN IS LESS MORE ? ATTENTION AND WORKLOAD IN AUDITORY , VISUAL , AND

REDUNDANT PATIENT-MONITORING CONDITIONS. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society*.

SHAMI, N.S., LESHED, G. AND KLEIN, D., 2005. Context of use evaluation of peripheral displays (CUEPD). In *INTERACT'05: Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*. Berlin, Heidelberg: Springer-Verlag, pp. 579–587.

SPARROW, B., LIU, J. AND WEGNER, D.M., 2011. Google effects on memory: cognitive consequences of having information at our fingertips. *Science*, 333, pp.776–778.

SRINIVASAN, N., SRIVASTAVA, P., LOHANI, M. AND BAIJAL, S., 2009. Focused and distributed attention. *Progress in Brain Research*, 176, pp.87–100.

TAKAHASHI, Y., KOJIMA, H. AND OKADA, K., 2011. Injured person information management during second triage. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. New York, New York, USA: ACM Press, p. 97.

THOMASSEN, O., BRATTEBØ, G., HELTNE, J.-K., SØFTELAND, E. AND ESPELAND, A., 2010. Checklists in the operating room: Help or hurdle? A qualitative study on health workers' experiences. *BMC Health Services Research*, 10(1), p.342.

TOVÉE, M.J., 2008. *An Introduction to the Visual System*, Cambridge University Press.

TUFTE, E.R., 1990. *Envisioning Information*, Cheshire, CT, USA: Graphics Press.

UNGERLEIDER, L.G. AND MISHKIN, M., 1982. Two visual systems. In G. E. Schneider, ed. *Analysis of Visual Behavior*. Cambridge, MA: MIT Press, pp. 549–586.

VERDAASDONK, E.G.G., STASSEN, L.P.S., WIDHIASMARA, P.P. AND DANKELMAN, J., 2009. Requirements for the design and implementation of checklists for surgical processes. *Surgical Endoscopy*, 23(4), pp.715–726.

VINCENT, C., NEALE, G. AND WOLOSHYNOWYCH, M., 2000. Adverse events in British hospitals: Preliminary retrospective record review. *British Medical Journal*, 322(7285), pp.517–519.

VOGEL, D. AND BALAKRISHNAN, R., 2004. Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users. *UIST '04*, 6(2), pp.137–146.

WALLACE, J.R., SCOTT, S.D., STUTZ, T., ENNS, T. AND INKPEN, K., 2009. Investigating teamwork and taskwork in single- and multi-display groupware systems. *Personal and Ubiquitous Computing*, 13(8).

WEAR, D., KOKINOVA, M., KECK-MCNULTY, C. AND AULTMAN, J., 2005. Pimping: perspectives of 4th year medical students. *Teaching and Learning in Medicine*, 17(2), pp.184–91.

WICKENS, C.D. AND MCCARLEY, J.S., 2007. *Applied Attention Theory*, CRC Press.

WINTERS, B.D., GURSES, A.P., LEHMANN, H., SEXTON, J.B., RAMPERSAD, C.J. AND PRONOVOST, P.J., 2009. Clinical review: checklists - translating evidence into practice. *Critical Care*, 13(6), p.210.

ZIEWACZ, J.E. ET AL., 2011. Crisis checklists for the operating room: development and pilot testing. *Journal of the American College of Surgeons*, 213(2), pp.212–217.e10.