Author's Accepted Manuscript

How do interruptions affect clinician performance in healthcare? Negotiating fidelity, control, and potential generalizability in the search for answers

Penelope M. Sanderson, Tobias Grundgeiger

International Journal of

Human-Computer Studies

Editors-in-Chief
E. Motta
S. Wiedenbeck

www.elsevier.com/locate/ijhcs

Cite this article as: Penelope M. Sanderson, Tobias Grundgeiger, How do interruptions affect clinician performance in healthcare? Negotiating fidelity, control, and potential generalizability in the search for answers, *Int. J. Human-Computer Studies,* http://dx.doi.org/10.1016/j.ijhcs.2014.11.003

1

How do interruptions affect clinician performance in healthcare?

Negotiating fidelity, control, and potential generalizability in the search for answers

Penelope M. Sanderson[a*] & Tobias Grundgeiger[b]

[a]School of Psychology, The University of Queensland, Brisbane, Australia

[b]Institut Human-Computer-Media, Julius-Maximilians-Universität Würzburg, Germany,

*Corresponding author. Penelope M. Sanderson, School of Psychology, The University of

Queensland, St Lucia, QLD 4072, Australia, Tel.: +61 7 3346 9529

E-mail address: psanderson@itee.uq.edu.au

Highlights

- Interruptions in healthcare are studied in the field, simulators, and laboratory.

- The goal is to test whether interruptions can harm clinicians and their patients.

- Methods used vary in fidelity, control exercised, and potential generalizability.

- Studies with low fidelity should be more representative of healthcare work.

- Research would be improved by programs of studies and improving individual studies.

3

Abstract

Interruptions and distractions are a feature of work in most complex sociotechnical systems in which people must handle multiple threads of work. Over the last 10-15 years there has been a crescendo of investigations and reviews into the question of the impact that interruptions and distractions have on safety-critical aspects of healthcare work such as medication administration, but findings are still inconclusive. Despite this, many healthcare communities have taken steps to reduce interruptions and distractions in safety-critical work tasks, a step that will usually do no harm but that may have unintended consequences. Investigations with a higher yield of certainty would provide better evidence and better guidance to healthcare communities. In this viewpoint paper we survey some key papers reporting investigations of interruptions and distractions in the field, in simulators, and in the laboratory. We also survey reports of field interventions to minimise interruptions and distractions with a view to improving the safety of medication administration. To analyse the papers adopting each form of investigation, we use the dimensions of fidelity, formal control exercised, and the potential generalizability to the field. We argue that studies of interruptions and distractions outside the healthcare clinical context, but intended to generalize to it, should become more formally representative of the cognitive context of healthcare work. Research would be improved if investigators would undertake programs of studies that successively achieve fidelity, control, and potential generalizability, or would take the opportunity to improve the design of individual studies.

Keywords: Interruptions, distractions, multitasking, healthcare, representative design.

1.   Introduction

1.1.   The problem

Work in complex sociotechnical systems is usually multiply-threaded. There have been many treatments of this issue in aviation, process control, and air traffic control (Colom, Martinez-Molina, Shih, & Santacreu, 2010; Loukopoulos, Dismukes, & Barshi, 2009; Mumaw, Roth, Vicente, & Burns, 2000; Wickens, 2002). Over the last 10-15 years, concerns about some of the consequences of multiply-threaded work have emerged in the healthcare domain. Specifically, there has been a crescendo of empirical research as well as literature reviews on workplace interruptions and distractions in healthcare.

There are two main factors driving these concerns. First, the interest stems from healthcare workers' subjective responses to the interruptions and distractions they experience, including an increase in subjective workload and a sense of frustration. Second, there is the concern that interruptions and distractions may lead to errors in the performance of healthcare tasks, which may in turn cause harm to patients. For both reasons, researchers and practitioners have sought (1) to uncover the burden of the problem of interruptions and distractions in healthcare, and (2) to design and evaluate interventions to reduce the burden.

1.2.   Goal of paper

Our goal in this viewpoint paper is to survey the methods that researchers have used to study interruptions and distractions in healthcare, highlight cases of exceptionally good practice, and reflect on how empirical investigations might deliver more value with respect to (1) and (2) above. We are not attempting an exhaustive review and methodological classification of all investigations in the area, but instead we have selected important and influential studies that help us to illustrate the points we wish to make.

2.   Interruptions in healthcare

2.1.   Definitions of interruptions and distractions

Up to this point we have used the phrase "interruptions and distractions" to characterise the topic of this paper, because most of the healthcare literature refers to "interruptions and distractions". Within healthcare research there is some use of the term multitasking (Chisholm, Collison, Nelson, & Cordell, 2000; Laxmisan et al., 2007; Walter, Li, Dunsmuir, & Westbrook, 2013) where it tends to refer to the clinician's management of, and switching between, multiple threads of responsibility, rather than the clinician's timesharing or rapid switching between tasks at a molecular level. Using Salvucci, Taatgen, and Borst (2009) multitasking continuum, *sequential* multitasking and switching from one responsibility to another is usually the concern in healthcare (Walter et al., 2013), rather than *concurrent* multitasking. Sequential multitasking is more likely to be recorded as task switching in response to a series of interruptions. Concurrent multitasking at the most molecular level is usually not a favoured strategy for handling multiple threads of responsibility, given the safety-critical nature of healthcare tasks, unless cognitive resource demands make it possible (Wickens, 2002). Concurrent multitasking is often recorded as a distraction.

A further concern is that the terms "interruption" and "distraction" cannot refer a priori to certain classes of external events, because both terms require observation of a person's reaction before they can have meaning. Under most definitions (see below), requesting a person's attention (via a vocal request, via equipment alarm, via phone, via personal proximity) becomes an interruption only if the person ceases activity on their current task for a measurable amount of time. Similarly, a noisy background conversation or event becomes a distraction only if there is a measurable effect on a person's performance.

Within the healthcare literature there has been considerable variation in how interruptions and distractions are defined and how they are distinguished operationally during empirical investigations (for some examples of differences in definitions, see Grundgeiger & Sanderson,

2009; Sasangohar, Donmez, Trbovich, & Easty, 2012). For present purposes, and as implied above, we say that an interruption occurs when an event leads a person to remove their attention fully but temporarily from a primary, or current, task to another task, and then move their attention back to the primary task. An example is an intensive care nurse suspending a patient assessment while countersigning a medication order. We say that a *distraction* occurs when a person's attention is partially diverted from a primary task to another task but performance on the primary task is not fully suspended. An example is responding vocally to questions while performing a manual medical procedure. If the other task is sustained, we may talk of *multitasking*. Note that the definitions do not take into account the content, convenience, and usefulness of the two tasks. In the extreme, clinicians may not even consider events such as those described above as interruptions or distractions, because their content progresses clinical work.

Most of the research on interruptions and distractions in healthcare has been performed with doctors or nurses as participants. In what follows, when referring to healthcare participants in general we will use the term "clinicians" to cover both disciplines. By "clinicians" we refer to the fact that the doctors and nurses are working in a clinical context, which is usually a hospital.

2.2. Forms of investigation

Three key motivations for investigating interruptions and distractions in healthcare are to determine the burden they pose on clinicians, to identify whether and when they cause harm to patients, and to test interventions intended to reduce any such harm. Investigations that are informative for healthcare have generally taken one of four forms: (1) field investigations, (2) simulator-based investigations, (3) laboratory-based investigations, and (4) intervention studies. In this section we provide a brief overview of these general forms of investigation before introducing the conceptual framework that we will use to highlight methodological issues.

Field investigations take place in clinical contexts with clinicians as participants. They can have an ethnographic motivation (Colligan & Bass, 2012; Rivera, 2014), they can be focused on identifying and classifying activity (Berg et al., 2013; Trbovich et al., 2013; Weigl, Müller,

Zupanc, Glaser, & Angerer, 2011; Westbrook, Woods, Rob, Dunsmuir, & Day, 2010) or they can require clinicians to keep a diary (Baethge & Rigotti, 2013). A frequent motivation underlying field studies has been to identify the burden that interruptions and distractions impose on clinicians by collecting information on how often and under what conditions they occur. More rarely, field investigators collect information on the motivations of interrupters (Rivera, 2014) on the correctness of clinical procedures and on episodes of actual or potential harm, and they seek associations between interruptions and distractions and non-nominal behaviour or events (Westbrook et al., 2010).

Simulator-based investigations take place outside the context of delivering care to live patients. They help investigators clarify the conditions under which interruptions may or may not produce harm. Simulator-based investigations may be mounted in a full-scale healthcare simulation environment (Feuerbacher, Funk, Spight, Diggs, & Hunter, 2012; Liu, Grundgeiger, Sanderson, Jenkins, & Leane, 2009; Prakash et al., 2014) or in a part-task simulation environment (Magrabi, Li, Day, & Coiera, 2010). They typically involve clinicians as participants. As a form of investigation, simulator-based investigations show greater variety than either the field or laboratory-based investigation because they loosen the constraints both of the field and of the laboratory. By offering the opportunity for control in a safe environment, they not only help investigators clarify the conditions under which interruptions might produce harm, but also offer the opportunity to test interventions that might reduce harm.

In contrast to both field and simulator-based investigations, laboratory-based investigations involving interruptions have generally not been motivated by the practical problem of interruptions in healthcare, although investigators sometimes make claims about the potential generalizability of their results to such problems (Monk, Trafton, & Boehm-Davis, 2008). Instead, laboratory-based investigations are generally performed to develop and test cognitive theories and models relating to memory and attention (for example, Altmann & Trafton, 2002; Dismukes & Nowinski, 2007). In laboratory experiments, factors such as the exact time of arrival of an

interruption, its duration, any advance warning of the interruption, the availability of visual cues relating to the original task, and so on, have been manipulated to distinguish different theories and build effective models. Nonetheless, some laboratory tasks offer findings that can be useful for healthcare if a case can be made for the generalizability of the findings.

To date, most intervention studies relating to interruptions in healthcare have taken place in the field, but field interventions can also be supplemented by trial interventions in a simulator context or even a laboratory context, in preliminary evaluations of effectiveness. Rather than seeking to establish relationships between interruptions and distractions and patterns of work in the field, intervention studies test the effectiveness of a workplace design (a novel workplace practice or device) that represents a hypothesis about how work practice and outcomes might be improved in a certain work context (Woods, 2003)

3. FCG cube

In this section we introduce the conceptual framework we will use to discuss methodological aspects of present research on interruptions and distractions in healthcare. As Brinberg and McGrath (1985) and Woods (1985) have noted, any behavioural investigation has a degree of fidelity (apparent realism in relation to practice in a substantive domain), control (specificity of inference and precision of measurement), and generalizability (potential for depth of insight and scope of application of conclusions, often driven by theory). In operationalizing those three concepts for present purposes, we make use of the concept of "representative design" (Brunswik, 1955, 1956; Goldstein, 2006; Hammond & Stewart, 2001). Just as participants for an experiment are normally sampled from the population to which we want to generalize, representativeness is the degree to which the stimuli and conditions used in an experiment have been intentionally sampled to reflect the range of contexts to which investigators intend that conclusions should apply.

3.1. Fidelity

Fidelity is the apparent realism of the investigative context in relation to the domain itself. For the study of interruptions in healthcare, therefore, fidelity refers to how tightly activity in an investigative context resembles activity in the healthcare context.

As many researchers have noted, fidelity is not a function solely of the physical attributes of the investigative context—in other words, it is not a function of how much the investigative context looks like the domain context (Dieckmann, Manser, Wehner, & Rall, 2007). Instead, it is much more closely related to the participant's experience. For our purposes, fidelity is high only if the professional *competence* of participants, the *situations* that they experience, and the *tools* available to them together allow them to perform according to the values and standards of their professional practice (Cumin & Merry, 2007; Woods, 2003; Woods & Christoffersen, 2002). Only then can participants enter into the social and emotional frame of professional behaviour (Dieckmann, Gaba, & Rall, 2007).

Fidelity depends on participants having a level of competence that makes them legitimate practitioners in the domain of interest. They do not need to be unusually expert in the domain. Furthermore, situations do not necessarily need to be fully-featured or tools complete for a scenario to achieve good fidelity within a clearly specified range.

When describing the studies included in this overview, we will evaluate their fidelity with respect to healthcare, taking into account the competence, situations, and tools provided in the study. Note that an investigative context might have good fidelity with respect to the domain itself, but may not represent the specific situations, tools, and levels of competence to which the investigator wishes to generalise. Good fidelity does not guarantee good generalizability; therefore generalizability is handled separately.

3.2. Control

Control refers to the measures taken to ensure that the conclusions of an investigation are specific and logically defensible. Control usually refers to the degree of precision with which the

investigator selects and manipulates participants, situations, and tools in order to identify the cause or causes of the findings (D'Amato, 1970). Manipulation or selection will ensure that properties of participants, situations, and tools that are actually or potentially relevant for performance are either held constant or varied in a known manner. Control sets a limit on the internal validity of an investigation, which in turn sets a limit on its maximum achievable external validity (Shadish, Cook, & Campbell, 2002).

When describing the studies included in this review, we will evaluate their level of control with respect to such factors as selection of participants and assignment to conditions, presence of controlled contrasts, standardisation of situations, reactive or non-reactive experimental arrangements, and so on. For present purposes we distinguish control from forms of analytic rigor found in ethnographic or interpretive research, such as methods for extracting themes or relationships.

3.3. Potential generalizability

Generalizability refers to the potential for depth of insight and breadth of application of conclusions. In our analysis we will refer to *potential* generalizability, which will indicate whether investigators have provided an *a priori* formal basis from which others can extend their conclusions (rather than whether claims have simply been made for greater applicability without further substantiation).

Generalizability of findings stems from (1) the groundedness of conclusions in abstract principles or theory and (2) a statement of the scope of the conclusions plus a specification of the pathway by which the conclusions would apply in the future. The more faithfully the investigative context distils the competence, situations, and tools that are relevant for practice, and represents the variables and relationships at play, the more certainty there is about the scope of potential generalizability(Brunswik, 1956; Hammond & Wascoe, 1980; Kirlik, 2006). In addition, the more that the competence, situations, and tools, variables and relationships can be described in abstract or theoretical terms, the more a pathway for applying the conclusions has been specified.

Note that a laboratory study testing a general theory with a high level of control, but with low representativeness with respect to targeted healthcare contexts, might have the internal validity to draw a conclusion such as "interruptions cause errors". However, the conclusion would have low potential generalizability unless the key work demands and constraints in the targeted healthcare context had previously been investigated and explicitly represented in the demands and constraints of the laboratory tasks.

Accordingly, when describing the studies included in this review, we will provide an approximate evaluation of their level of potential generalizability with respect to the criteria listed in the paragraph above: (1) groundedness in abstract principles or theory, and (2) stated scope and pathway for generalisation.

4.   Review of studies

In our review of studies we use the dimensions of fidelity, control, and potential generalizability to discuss the strengths and weaknesses of different forms of investigation for understanding the effect of interruptions and distractions on healthcare work processes and outcomes. We are not intending to provide an exhaustive review of the literature. Instead we wish to show how investigations in field, simulator, and laboratory contexts as well as intervention studies, have handled fidelity, formal control, and potential generalizability, and to indicate where research might be improved. A similar analysis could be done for any other field in which interruptions and distractions are a concern.

Table 1 lists the studies we have selected that illustrate of each form of investigation. The studies selected are all strong representatives of their form of investigation, but have also been selected to show some of the variety within that form of investigation. Some studies are well-established and highly-cited; others are very recent or less well known, but are strong examples of the strengths and challenges of their form of investigation, providing balance to the coverage.

In Table 1 we have provided a brief description of each study. Then we have commented on the fidelity, control, and potential generalizability of the study. Adjacent to each dimension for

each study is a small linear graphic that provides an approximate placement of the study on the dimension in question. Needless to say, the placement is approximate, and refers only to our judgment with respect to how we define each dimension for the purposes of the present paper. Our goal is to demonstrate general patterns of research, rather than to target specific papers. As noted, we have deliberately selected papers that are strong representatives of their type.

Figure 1 represents the three dimensions as the axes of a cube and positions each study in the three-dimensional space. The top right hand corner represents a combination of high fidelity, high control, and high potential generalizability that is probably unattainable in a single experiment. Once populated with studies, Figure 1 allows systematic similarities and differences to emerge in a graphic form between studies that use each form of investigation. It also reveals gaps between forms of investigation, indicating constraints in the investigative practices observed that could be overcome, so offering a greater insight.

4.1. Field studies

There is a vast number of field investigations of interruptions and distractions in healthcare (see reviews by Coiera, 2012; Grundgeiger & Sanderson, 2009; Hopkinson & Jennings, 2013; Li, Magrabi, & Coiera, 2012; Rivera & Karsh, 2010). We have selected three to discuss: Westbrook et al.'s (2010) observational study of the impact of interruptions on nurses' work during medication rounds, Grundgeiger, Sanderson, MacDougall, and Venkatesh's (2010) eye-tracking study of nurses' ability to resume interrupted tasks, and Rivera's (2014) field study of how nurses make decisions to interrupt other nurses, combining ethnographic observation and interviews.

Fidelity. The observational arms of all three studies are moderately high to high in fidelity, involving professionally competent practitioners working in representative situations with their normal tools. Deviations were the presence of an observer with coding tablet or notebook in the Westbrook et al. (2010) and Rivera (2014) cases, and the fact that a nurse wore the eyetracker in the Grundgeiger et al. (2010) case. The interview arm of Rivera's study has less fidelity, however, as it required participants to make abstractions from their experience to answer questions about

when, why, and how nurses interrupt each other and what the consequences are.

Control. Formal control of field studies is generally low. All three studies constrained the selection of healthcare contexts in which to study interruptions to some degree, with Westbrook et al. (2010) choosing medication administration in general wards across two hospitals, Grundgeiger et al. (2010) choosing the first three hours of the bedside ICU nurse's shift, and Rivera (2014) choosing nursing work in a neuroscience surgical ICU.

Potential generalizability. Despite the above similarities, the purposes of the three papers were very different and therefore the nature of their potential generalizability varies considerably. In their observational study, Westbrook et al. (2010) sought an association between the number of interruptions a nurse received while administering medication, and the number of procedural failures and clinical errors the nurse exhibited in the same medication administration round. By noting that an association between the number of interruptions and the number of procedural failures/clinical errors was found in both hospitals, and by characterising the association as a "dose-response relationship", Westbrook et al. invoked a statistical biomedical concept to indicate potential generalizability. However no explanatory model was provided; it is solely a statistical association. Westbrook et al. (2010) discuss the limits to potential generalizability of their study, such as the lack of sampling of medication administration at times other than day shifts. However they do not provide a theoretical basis for generalising the findings to other kinds of activities, either within nursing, within healthcare, or within safety-critical collaborative work. Indeed, Westbrook et al. call for further research that helps us understand why interruptions occur, how they are managed, and how staff judge when to interrupt. As we will see, these are questions investigated by Rivera (2014).

In a contrast to Westbrook's approach, Grundgeiger et al. (2010) used pre-existing theory to motivate their analysis of impact of interruptions on nursing work. They drew upon the memory for goals theory (Altman & Trafton, 2002) and the associative activation model (Dismukes & Nowinski, 2007) to identify prospectively six factors that might influence how long

it would take nurses to resume their primary task after an interruption (the "resumption lag").

Regression analyses showed that two factors—the length of the interruption and the presence of a

change in work context during the interruption—were positively associated with longer

resumption lags. However, the above regression analysis could be applied only to about half of the

interruptions observed because nurses' strategies for handling multiple work threads often

removed the resumption lag altogether. Theoretical explanations for the nurses' strategies were

suggested retrospectively, but their generality remains untested.

Finally, Rivera's (2014) ethnographic study of interruptions in an NSICU starts to provide

an answer to Westbrook et al. (2010) question of why nurses interrupt other clinicians, but its

potential generalizability is not clear, both because it was the author's first investigative study in

the NSICU and because of its ethnographic stance. In her analysis, Rivera identified and classified

behaviour-shaping constraints relating to the work context of the NSICU, such as the size of the

unit, the nature of ICU work, communication norms, and so on. However the abstractions

achieved were not generalised outside the NSICU context. Similarly, factors shaping nurses'

perceptions of the "interruptibility" of other nurses and of the kinds of interruptions that may be

warranted were outlined, such as the interrupter's experience, patient consequences, and so on, but

whether and how those factors would generalise was not discussed. Rivera (2014) acknowledges

the limitations of having used one setting only and notes that further research is needed. She

suggests that the ability of other researchers to judge the "transferability" of her findings at this

point rests principally in the detail provided about the setting, context, analysis, and findings of

the NSICU. This is in contrast to the use of abstract principles derived from the data or the use of

theory.

4.2. Simulation studies

Technical developments and the increased use of healthcare simulation facilities for

training have made it possible to use medical simulators for research as well (Merry et al., 2008).

The apparent benefits of using simulators to study interruptions in healthcare are the opportunity

to increase fidelity of participants, situations, and tools, and the opportunity to exercise formal experimental control without ethical issues or organizational concerns. We highlight three simulation studies: (1) Feuerbacher et al.'s (2012) test of whether operating room distractions and interruptions would induce errors by novice surgeons, (2) Magrabi et al.'s (2010) test of whether the absence vs. presence of interruptions and task complexity would affect whether physicians make errors in prescribing medication using a computerized provider order entry system, and (3) Prakash et al. (2014) study of whether interruptions cause oncology nurses to notice fewer errors during medication verification and commit more errors during medication administration.

Fidelity. All three studies included competent subject matter experts and seemed to provide all necessary tools for the tasks (see Table 1). In relation to situations, in the Feuerbacher et al. (2012) and the Magrabi et al. (2010) studies, interruptions were operationalized as the observer or experimenter intervening in the simulated scenario or task. We argue that interventions by the observer may interfere with the fidelity of the study, because the represented situation is not what a clinician would experience. In fact, "fiction cues"—cues that emphasize the artificial/simulated situation—will change the experience of participants and may change the behavior of the participants (Dieckmann et al., 2007b). In addition, participants in simulations may vary in their perception of which cues or events are part of the scenario vs. part of reality outside the scenario, which can compromise control (Dieckmann et al., 2007a; 2007b). In addition, the simulated situation becomes less representative in studies that constrain subject matter experts in the kind of interruption management strategies they can use (Magrabi et al., 2010). The Prakash et al. (2014) study solves the above issues by using naturalistic scenarios and introducing interruptions as actions by other actors, such as telephone calls or requests from other nurses or from patients (all actors) to distract or interrupt, thereby using content that is consistent with the presented scenarios (Grundgeiger et al., 2013; Liu et al., 2009).

Control. In relation to control, the studies selected either included an experimental contrast between participants and used scripted procedures (Feuerbacher et al., 2012; Prakash et al., 2014)

or conducted a detailed analysis of the tasks and mounted a within-participants design (Magrabi et al., 2010). To insure internal validity, within-participant designs should report how potential order effects are handled. Such an analysis, however, is difficult if the total number of participants and the number of observation per participants are low (Grundgeiger et al., 2013; Liu et al., 2009; Merry et al., 2008). Between-participants manipulations provide stronger conclusions.

Potential Generalizability. The potential generalizability of findings from simulator studies to interruptions in the field may be high if abstract principles or theories are invoked, and if the results allow properties that contribute to the incidence and impact of interruptions to be related directly and convincingly to the abstract principles and theories that are invoked. Only one of our simulation examples uses an approach that borrows such principles. In their study of doctors using a CPOE system, Magrabi et al. (2010) manipulated factors derived from prior laboratory-based interruptions that would be expected to affect the impact of interruptions. As noted, however, the study has restrictions on representativeness that limit its generalizability.

In the Feuerbacher et al. (2012) simulator study of operating room interruptions and distractions which included virtual reality, the specific interruptions and distractions used in the study were sampled from prior observations. Generalization was based on the typicality or realism of the sampled interruptions for the surgical context rather than on a theory of why they those interruptions and distractions might affect surgical performance. The study's representativeness limitations are less constraining compared to those of Magrabi et al. (2010) in the sense there was greater freedom of action for participants, but there is little basis in abstract principles or theory for predicting the effect of further interruptions and distractions that were not included in the study. Finally, in the Prakash et al. (2014) study no recourse was made to theory and no overt effort made to generalize the findings outside the oncology unit. The above studies underscore the need for more theoretically-guided simulator studies that will provide stronger bases for generalization.

4.3.   Laboratory studies

The effects of interruptions on humans have been studied extensively in the laboratory (Li, Blandford, Cairns, & Young, 2008; Trafton, Altmann, Brock, & Mintz, 2003; for a summary see Trafton & Monk, 2007). Many laboratory studies of interruptions measure the effects of theoretically-motivated manipulations on highly sensitive aspects of human performance, such as differences in latency in the range of milliseconds or seconds.

As is well known, laboratory research places a strong emphasis on experimental control and internal validity. Frequently this emphasis results in compromised fidelity and, because the actual intended area of application of the result is not well defined, uncertain representativeness. For the present purpose, we consider interruptions in healthcare as one broad intended area of application of laboratory research on interruptions—indeed, healthcare is often mentioned as a potential area of application by laboratory investigators (see for example Monk et al., 2007).

The laboratory studies we have selected for discussion are (1) Bogunovich and Salvucci's (2011) investigation of how participants manage deferrable interruptions with small vs. large time constraints, (2) Brumby, Cox, Back, and Gould (2013) study of the effect of interruption lags on resumption errors using a donut-making microworld, and (3) Cao and Liu (2013) study of diagnostic judgment accompanied by monitoring and memorisation multitasking demands. We have chosen these studies in part because they measured observable memory effects such as forgetting or diagnostic judgments rather than latency alone. Forgetting and judgment have greater potential consequences in healthcare than a few seconds' difference in responding.

Fidelity All three experiments used student participants rather than participants who were competent practitioners. In both the Bogunovich and Salvucci (2011) and Brumby et al. (2013) studies, task content was unrelated to healthcare. The experimental tasks were rather simple computer-based tasks, and there was no specification of what kind of tasks in the field the experimental tasks might represent. In contrast, the Cao and Liu (2013) study specifically addressed diagnostic decision-making in healthcare under multitasking and interruptive

conditions. However, the diagnostic task used in the study was highly simplified and no case was made that the multitasking imposed would be experienced in healthcare in the manner presented.

In both Brumby et al. (2013) and Cao and Liu (2013) studies, participants were not allowed any discretionary control on how they could manage interruptions, but instead were abruptly interrupted and had to resume at a specific point without the option of restarting a subtask.

In contrast, Bogunovich and Salvucci's (2011) student participants worked on a computer and were interrupted by a ringing phone. Their experimental set-up may have had modest similarities to aspects of clinical work and the participants were probably familiar with such a task. Importantly, participants were given discretionary control on exactly when they could answer the phone. Discretionary control of interruption management has been shown to be an important factor in healthcare (Colligan & Bass, 2012; Grundgeiger et al., 2010; Liu et al., 2009) but laboratory studies frequently constrain participants' behaviour. Investigators performing laboratory-based research that is intended to generalize to healthcare tasks should conduct a cognitive task analysis (Hoffman & Militello, 2008) of the healthcare tasks to ensure that properties of the healthcare task that are likely to be affected by interruptions have structural analogs in the laboratory task.

Control. With regard to control, all three studies used an experimental design and exerted tight control on task selection and on the timing of the task steps and interruptions. Accordingly, internal validity is high.

Potential Generalizability. The present examples were rated low or moderate for their potential generalizability. Brumby et al. (2013) and Cao and Liu (2013) explicitly invoke abstract principles or theory but they do not fully address the issue of the representativeness of their experimental arrangements for other contexts. Even though Brumby et al. (2013) used an established theory (memory for goals, Trafton et al., 2003), they did not specify the kind of work tasks for which the results may be relevant and did not ensure that the laboratory task reflected the

structural properties and cognitive demands of any particular class of work tasks.

In contrast, Cao and Liu (2013) specifically targeted medical diagnostic decision-making. However, the structure of the diagnostic problem space they used as the experimental task, and the relationship of the timeshared tasks to the diagnosis task, were not unambiguously distilled from the domain itself in a way that would make generalization to the domain straightforward. When discussing future work, Cao and Liu proposed capturing more of the competence, situations, tools, and domain complexities of medical diagnostic decision-making. In this way they will remove levels of control and task abstraction that threaten representativeness and potential generalizability.

Bogunovich and Salvucci (2011) captured a little more of the discretion that workers, including clinicians, can exercise in when handling interruptions. As a result, rather than solely testing the impact of current workload on whether participants accept a phone call, Bogunovich and Salvucci were able to identify time constraints and number of steps to the next low-workload point as further factors. Although the latter two abstractions are post-hoc, they are properties that account for some aspects of interruption management that could be applied to other situations.

4.4. Intervention studies

We present the intervention studies separately because their authors intend to *change* practice rather than to describe or explain practice. A recent review of studies that have tested interventions to reduce interruptions and, by implication, to reduce medication administration errors, has noted that there is only weak evidence that such interventions are effective (Raban & Westbrook, 2014). In this section we highlight a "multi-intervention" study by Tomietto, Sartor, Mazzocoli, and Palese (2012), which led to conflicting outcomes, a medication administration accuracy study by Kliger, Singer, Hoffman, and O'Neil (2012) that included minimising interruptions and distractions, and an early study of interventions by Pape (2003) to reduce interruptions. Interventions are sometimes the end-point of a sustained program of research into interruptions (Colligan, Guerlain, Steck, & Hoke, 2012; Prakash et al., 2014; Trbovich, Howell, et

al., 2010) and programmatic research will be covered in a later section.

Fidelity. All the studies in this category have high to moderately-high fidelity. They were carried out in the field using competent practitioners as participants, and started with work situations that are habitual and important and that include the participant's normal work resources and tools. The interventions intended to reduce interruptions usually consisted of changes to work situations (for example, reductions in how often the participant's attention may be called to other tasks) and changes in work resources and tools (for example, changes in where tasks are done, introduction of new checklists or devices supporting tasks).

Deviations from the "normal" frame of healthcare work can arise from constraints associated with collecting data in field contexts. In some cases the constraints might compromise the plausibility or sustainability of the intervention. The study by Pape (2003) in which a single medical-surgical nursing unit experienced a control (baseline) period, a first ("focused") intervention period and then a second ("Medsafe vest") intervention period in close succession may have been such a case.

Control. Intervention studies have logistical and organisational challenges that make it operationally difficult, or ethically unacceptable, to exercise formal control. Accordingly we give the issue of control a fuller treatment for the intervention studies than for the other kinds of studies.

As noted by Raban and Westbrook (2014), almost all intervention studies (with the exception of Pape, 2003) have been quasi-experiments with a pre-post design. Assigning participants at random to conditions that are run in parallel with each other is impractical, due to the need to keep work practices consistent within hospital units. Assigning hospital units at random to conditions risks introducing confounds, due to other differences between the units and due to the difficulties of keeping treatments independent within a highly integrated organisation such as a hospital. Assigning different hospitals to conditions may reduce the problem of keeping treatments independent, but exacerbates the potential for confounds. Creating formal contrasts is

therefore challenging.

In professional contexts, orthogonal comparisons of the effects of different interventions may be too expensive, so the different interventions are often introduced at the same time. In their large, broad-based quality improvement project across six hospitals over several years, Kliger et al. (2012) introduced six safety processes to improve the accuracy of medication administration that included the goal of protecting the process from distractions and interruptions. Because of the multiple interventions, any improvement found in the accuracy of medication administration could not be attributed uniquely to reductions in interruptions and distractions. Further, different combinations of interventions to reduce interruptions were adopted in different hospital units. Even if the reduction in interruptions had shown stronger associations with better medication administration than for the other interventions, it would still be unclear whether all means of reducing interruptions are equally effective or whether some are more effective than others.

The use of combinations of interventions such as those in Kliger et al. (2012) can leave paradoxes unresolved. In the Tomietto et al. (2012) study, the interventions intended to reduce interruptions included a special medication preparation room, a red tabard to be worn by the nurse doing the medication round, and general education of staff on the new changes. Although the total number of interruptions decreased after the above interventions were introduced, Tomietto et al found that interruptions by staff to the nurse actually become *more* frequent, but shorter, whereas interruptions by patients to the nurse became less frequent and shorter. It is unclear which part of the intervention, or all parts, led to the unexpected change in the pattern of staff interruptions. These findings point to a failure to capture underlying motivations for interruptions.

A further control issue relates to the period over which observations are made. Kliger et al.'s (2012) pre-post study ranges over periods of years, during which time many other factors than the broad-based interventions included in the study may have come into play. In contrast, Pape (2003) reports one intervention introduced for just eight successive medication rounds, rapidly followed by an extended intervention introduced for a further eight successive medication

rounds. Order effects and a possible diffusion of the treatment effect compromise conclusions that can be drawn.

Potential generalizability. The three intervention studies described have not been motivated by theoretical accounts of how interruptions might lead to harm, but instead by the practical goal of improving the accuracy of medication administration and removing the potential for error. Indeed, given that the evidence for a causal connection between interruptions and errors is still tenuous (Grundgeiger & Sanderson, 2009; Hopkinson & Jennings, 2013; Raban & Westbrook, 2014) and given reports of paradoxical outcomes (Tomietto et al., 2012) it is arguable that intervention studies are premature. Elsewhere we have argued that to draw a connection between interruptions and harm, we need a theory not just of the effect of interruptions on human cognition, but also of how accidents occur (Grundgeiger & Sanderson, 2009).

A further constraint on the potential generalizability of intervention studies is that investigators have focused largely on reporting outcomes in their own area of practice and have not been required to think beyond that. The investigations have not offered systematic comparisons of the effectiveness of an intervention between hospitals or between areas of practice that might increase our confidence that the manipulations have general applicability. Nor do the investigations offer analyses of how the effectiveness of the manipulations might be conditioned by contextual variables.

5.   Towards greater potential generalizability

In this section we discuss how investigations into the role of interruptions and distractions in healthcare might improve their potential generalizability. First we discuss programmatic research that incorporates multiple studies, and we present further details of an example touched on earlier. Then we discuss how the design of individual studies might achieve better potential generalizability.

5.1.   Programmatic research

One way to achieve high generalizability of outcomes to practice is through a program of

research that uses multiple forms of investigation in series, successfully benefitting from the representativeness of the investigations, the use of theory, and the precision with which causal statements can be made. Investigations with high fidelity can stimulate ideas about how a phenomenon emerges in a situation of a concern that can then be connected with theory and tested in more controlled settings. If the controlled settings have been designed to be representative of the targeted healthcare settings, then the conclusions are likely to generalise to those settings.

Many researchers investigating interruptions and distractions tend to persist with one form of investigation. There is relatively little evidence in the field of programmatic research that traverses different forms of investigation. One exception is recent work of Trbovich and colleagues (Prakash et al., 2014; Trbovich et al., 2013; Trbovich, Howell, et al., 2010; Trbovich, Prakash, Stewart, Trip, & Savage, 2010). One phase of that work was described earlier, but here we present the broader program. The phases of the program are shown in Figure 1, linked together.

Under Canadian Patient Safety Institute funding, Trbovich and colleagues conducted a multiphase study into the effect of interruptions on medication administration and the potential for interventions to improve safety (Trbovich, Howell, et al., 2010). In Phase 1, performed in the field, they shadowed oncology nurses who were administering medications to patients, and gathered information about the sources and frequencies of interruptions (Trbovich, Prakash, et al., 2010) (see node P1 on Figure 1). They identified tasks that were most likely to be interrupted and they found that tasks generally took longer to complete when nurses were interrupted. In Phase 2, performed in a full-scale simulation environment, some of the situations observed in Phase 1 were simulated (see node 6 on Figure 1). A controlled and counterbalanced manipulation of interrupting vs. not interrupting participants was used (see Prakash et al. 2014 and the discussion of that study in the simulation section herein). Nurses made more medication administration errors when they were interrupted than when they were not interrupted; moreover, they were also less likely to notice errors "planted" in the scenario when interrupted

The initial field study and simulation study just described provided baseline data against which Prakash et al. (2014) could test the effectiveness of interventions in the simulator and the field. In Phase 3, the researchers worked with domain practitioners to conceptualise and develop interventions intended to reduce interruptions: medication verification booths, visual timers, motor sensor lamps, informative signage, vocalising task steps, and standardising workflow (not shown on Figure 1). In a simulation-based intervention study (Phase 4) the effectiveness of some of the interventions was tested using the same scenarios as in Phase 2 (see node P4 of Figure 1). After the intervention was introduced, nurses made fewer errors and were more likely to detect the planted errors. Finally, in an intervention field study (Phase 5) performed in an oncology centre, the researchers introduced the interventions that had previously been tested in the simulator and found that nurses experienced fewer interruptions during drug verification and pump programming tasks (Trbovich, Howell, et al., 2010) (see node P5 on Figure 1). Data were not available to evaluate whether there were fewer errors in medication administration with the interventions in place, so generalizability is not yet established. However, because such outcomes were found in the simulator on similar tasks and with similar interventions, a positive outcome is likely.

Alongside its obvious strengths, Trbovich's program has some shortcomings, in that opportunities were missed to run fully controlled simulator-based studies establishing cause-effect relationships between interruptions and errors, and between interventions and reductions in errors. The re-use of the "interrupted, no interventions" condition as a contrast for both the "uninterrupted" condition in the pre-intervention phase (Phase 2) and as a control for the "interrupted, with interventions" condition in the post-intervention phase (Phase 4) complicates interpretation. The rationale for the interventions is not strongly based in theory, although it is systematically and thoughtfully based in observation of practice to which findings would be generalised. Moreover, the representativeness of the simulator scenarios with respect a broader range clinical contexts was not formally analysed, so further potential generalizability to other

tasks, other forms of interruption, and other kinds of care contexts is not known. Finally, the potential for some interruptions to have positive effects and the potential for suppressing interruptions to create inconveniences for other clinicians has not been considered. Some of these themes will be amplified in later section. Despite these shortcomings, the program of Trbovich and colleagues is an excellent example of how a program of research using different forms of investigation can lead to a set of thoughtful and well-targeted interventions.

Not all researchers have access to the resources that supported the research program of Trbovich and colleagues. An important question is how researchers with fewer resources might still make contributions that have an impact. In our view, the answer lies in the deftness with which a researcher can address fidelity, control, and potential generalisability both within a study and across successive studies, while moving from a problem statement to the form of a solution. The ability to use different forms of investigation that offer the property that is most important for the present phase of an inquiry is critical, as are forming good relationships with "problem owners" or theoreticians. In what follows, we focus on how more might be made of specific forms of investigation in an individual study.

5.2.   Individual studies

Although it is unlikely that a single study can make the same contribution as an integrated research program, our survey has revealed opportunities to increase fidelity, control, and potential generalization. Specific forms of investigation (field, simulator, laboratory, and intervention) have specific weaknesses that we address in turn. Figure 1 shows gaps between clusters of studies sharing a form of investigation; the gaps indicate possible areas for stronger study design.

Field studies. The principal challenge for field studies is control. It is seldom possible to exert control by manipulation of treatments and random assignment because of ethical issues and organizational constraints. However, researchers may still achieve a high level of control by prospective theory-guided selection of field situations. For example, when observing interruptions and medication preparation error rates one can use modelling methods such as GOMS (Gray,

John, & Atwood, 1993) to distinguish between medication preparations that pose high vs. low

cognitive workload (for a similar apporach in a simulation see Magrabi et al., 2010). The contrast

between low and high workload could then be included in the analysis. By using a construct such

as workload, the potential generalizability of the study would be increased because the results

indicate a task characteristic that has a general property; using this procedure, workload could be

distinguished for every task and the idea tested that interruptions compromise performance over a

certain level of workload only. The same principle applies to other constructs.

Simulation studies. Simulations can offer a high level of fidelity. The challenge for

researchers who use simulations to study interruptions is to focus the fidelity towards situations

and tasks that are representative of where interruptions occur and where consequences of error are

high, and to add a level of control that removes competing explanations. In relation to fidelity, the

inherent competence of participants is not affected when control is exercised, but the tools

available to them and the situations they experience might be affected. In some of the simulation

studies discussed earlier, the researcher distracts or interrupts the participant and the interrupting

task may be quite arbitrary. This is not a representative situation in clinical work and it may

compromise the potential generalizability of results.

One way of overcoming artificialities is to select distractions and interruptions that are

thematically related to the "frame" of the study, and that arrive in a natural-appearing manner,

even if actually tightly controlled. In previous research on how ICU nurses remember future tasks

(prospective memory), Grundgeiger et al. (2013) consulted with subject-matter experts to

construct a 40-minute scenario of a start of a morning shift. The scenario included several

carefully selected and carefully timed distractions, such as a short 3-second vital sign alarm

sounding just as the participant was encoding a prospective memory task, or interruptions such as

a telephone call from the simulated patient's relative.

Furthermore, in simulation studies researchers have the opportunity to establish an

effective contrast between a non-interrupted baseline condition and an interrupted condition. To

avoid potential order or carryover effects, a between-participants manipulation should be preferred where resources allow.

Finally, it would encourage deeper thinking about the potential generalizability of findings, if reports of simulation studies were to regularly include details of scenario design, ongoing task characteristics, and interrupting task characteristics. The psychological literature on interruptions (Cumin & Merry, 2007; Trafton & Monk, 2007; Woods, 2003; Woods & Christoffersen, 2002), task switching (Monsell, 2003), or prospective memory (Dismukes, 2012; McDaniel & Einstein, 2000) can provide theoretical guidance that can inform scenario design or possible manipulations.

Laboratory studies. There is a long tradition of experimental laboratory studies being criticized for not sampling across situations (Brunswik, 1955) and for not studying phenomena as they occur in everyday life (Neisser, 1982). Laboratory studies would have greater applicability to healthcare if investigators systematically identified the cognitive and perceptual demands of tasks of interest in relation to interruptions. Based on an analysis of these demands, a laboratory task or microworld could be constructed that distils rather than dilutes key aspects of a selected healthcare context (Woods, 1985). The task or microworld will then let the intended participant use knowledge and judgment when working on the task, and will provide the required means or tools for engaging with the task, if relevant. Clearly, having a clinician participate in a task that restricts the use of the clinician's expertise, or in a task for which the clinician's expertise is not relevant, wastes an opportunity. Similarly, using a task that purports to be the task of an expert but conducting the study with student participants will remove the opportunity for high fidelity and representativeness. Thoughtfully constructed laboratory tasks can have high potential generalizability if their representativeness is as carefully engineered as the way they address theory.

Intervention studies. Intervention studies addressing interruptions and distractions in healthcare are largely directed at medication administration tasks and usually take place in the

field. As noted, the studies are based on the assumption that interruptions and distractions are harmful to work activities and that removing interruptions and distractions will improve work. Until there is better differentiation between interruptions and distractions that are helpful communication events promoting organizational resilience vs. those that are not helpful (Grundgeiger & Sanderson, 2009; Sasangohar et al., 2012) and better understand of the reasons that clinicians interrupt (Rivera, 2014), intervention studies will be difficult to design effectively and may lead to paradoxical results (Tomietto et al., 2012).

6.   Conclusion

In this paper we have used the concepts of fidelity, control, and potential generalizability to survey some representative papers addressing interruptions and distractions in healthcare, using different forms of investigation. Our goal has not been to provide an exhaustive review of the literature using these concepts. Instead, our goal has been to express a viewpoint on how the different forms of investigation are presently being used to address the issues of whether interruptions and distractions can disrupt clinicians' work to the point of causing harm, and whether interventions to reduce or remove interruptions and distractions improve clinicians' work and lessen the likelihood of harm.

Although we recognise that research investigations are usually limited in time and resources, making it difficult to mount studies that conform to ideal models, we encourage researchers to seek ways to achieve more generalizable results. As the contents of Figure 1 suggest, this may involve finding ways to increase the fidelity and representativeness of laboratory and simulation studies, increasing the ways that theories or abstractions proposed in other studies may guide control in any form of investigation, and increasing the degree of formal control actually exercised in simulation and field studies. Alternatively, achieving more generalizable results may involve planning investigations that use multiple forms of investigation that successively achieve the levels of representativeness and control required. Whichever path is

taken should improve the rate at which we arrive at an understanding of the role interruptions and

distractions play in healthcare, and it should improve the clarity of that understanding.

7.   References

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: an activation-based model. [Article].
Cognitive Science, 26(1), 39-83.

Baethge, A., & Rigotti, T. (2013). Interruptions to workflow: Their relationship with irritation and
satisfaction with performance, and the mediating roles of time pressure and mental
demands. [Article]. *Work and Stress, 27*(1), 43-63.

Berg, L. M., Kallberg, A. S., Goransson, K. E., Ostergren, J., Florin, J., & Ehrenberg, A. (2013).
Interruptions in emergency department work: An observational and interview study.
[Article]. *BMJ Quality & Safety, 22*(8), 656-663.

Bogunovich, P., & Salvucci, D. (2011). The effects of time constraints on user behavior for
deferrable interruptions. *Proceedings of the SIGCHI Conference on Human Factors in
Computing Systems*, 3123-3126.

Brinberg, D., & McGrath, J. E. (1985). *Validity and the research process*. California: Sage
Publications.

Brumby, D. P., Cox, A. L., Back, J., & Gould, S. J. J. (2013). Recovering from an interruption:
Investigating speed-accuracy trade-offs in task resumption behavior. *Journal of
Experimental Psychology-Applied, 19*(2), 95-107.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology.
Psychological Review, 62(3), 193-217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*.
Berkeley, CA: University of California Press.

Cao, S., & Liu, Y. (2013). Effects of concurrent tasks on diagnostic decision making: An
experimental investigation. *IIE Transactions on Healthcare Systems Engineering, 3*(4),
254-262.

Chisholm, C. D., Collison, E. K., Nelson, D. R., & Cordell, W. H. (2000). Emergency department
workplace interruptions: Are emergency physicians "interrupt-driven" and "multitasking"?

Academic Emergency Medicine, 7(11), 1239-1243.

Coiera, E. (2012). The science of interruption. *BMJ Quality & Safety, 21*(5), 357-360.

Colligan, L., & Bass, E. J. (2012). Interruption handling strategies during paediatric medication administration. *BMJ Quality & Safety, 21*(11), 912-917.

Colligan, L., Guerlain, S., Steck, S. E., & Hoke, T. R. (2012). Designing for distractions: a human factors approach to decreasing interruptions at a centralised medication station. *BMJ Quality & Safety, 21*(11), 939-947. doi: Doi 10.1136/Bmjqs-2011-000289

Colom, R., Martinez-Molina, A., Shih, P. C., & Santacreu, J. (2010). Intelligence, working memory, and multitasking performance. *Intelligence, 38*, 543-551.

Cumin, D., & Merry, A. (2007). Simulators for use in anaesthesia. *Anaesthesia, 62*(2), 151-162.

D'Amato, M. R. (1970). *Experimental psychology: Methodology, psychophysicis, and learning.* New York: McGraw-Hill.

Dieckmann, P., Gaba, D., & Rall, M. (2007). Deepening the theoretical foundations of patient simulation as social practice. *Simulation in Healthcare, 2*(3), 183-193.

Dieckmann, P., Manser, T., Wehner, T., & Rall, M. (2007). Reality and fiction cues in medical patient simulation: An interview study with anesthesiologists. *Journal of Cognitive Engineering and Decision Making, 1*(2), 148-168.

Dismukes, R. K. (2012). Prospective memory in workplace and everyday situations. *Current Directions in Psychological Science, 21*(4), 215-220.

Dismukes, R. K., & Nowinski, J. (2007). Prospective memory, concurrent task management, and pilot error. In A. Kramer, D. Wiegmann & A. Kirlik (Eds.), *Attention: From Theory to Practice* (pp. 225-236). New York: Oxford.

Feuerbacher, R. l., Funk, K. H., Spight, D. H., Diggs, B. S., & Hunter, J. G. (2012). Realistic distractions and interruptions that impair simulated surgical performance by novice surgeons. *Archives of Surgery, 147*(11), 1026-1030.

Goldstein, W. M. (2006). Introduction to Brunswikian theory and method. In A. Kirlik (Ed.),

Adaptive perspectives on human-technology interaction: Methods and models for cognitive engineering and human-computer interaction (pp. 10-24). Oxford; New York: Oxford University Press.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. *Human-Computer Interaction, 8*(3), 237-309.

Grundgeiger, T., Sanderson, P., MacDougall, H. G., & Venkatesh, B. (2010). Interruption Management in the Intensive Care Unit: Predicting Resumption Times and Assessing Distributed Support. *Journal of Experimental Psychology-Applied, 16*(4), 317-334.

Grundgeiger, T., & Sanderson, P. M. (2009). Interruptions in healthcare: Theoretical views. *International Journal of Medical Informatics, 78*(5), 293-307.

Grundgeiger, T., Sanderson, P. M., Beltran Orihuela, C., Thompson, A., MacDougall, H. G., Nunnink, L., & Venkatesh, B. (2013). Prospective memory in intensive care nursing: A representative and controlled patient simulator study. *Ergonomics, 56*(4), 579-589.

Hammond, K. R., & Stewart, T. R. (2001). *The essential Brunswik: Beginnings, explications, applications*. New York: NY: Oxford University Press.

Hammond, K. R., & Wascoe, N. E. (1980). *Realizations of Brunswik's representative design*: Jossey-Bass San Francisco, CA.

Hoffman, R., & Militello, L. (2008). *Perspectives on cognitive task analysis*. New York: Academic Press.

Hopkinson, S. G., & Jennings, B. M. (2013). Interruptions during nurses' work: A state of the science review. *Research in Nursing and Health, 36*(1), 38-53.

Kirlik, A. (Ed.). (2006). *Adaptive perspectives on human-technology interaction: Methods and models for cognitive engineering and human-computer interaction*. Oxford; New York: Oxford University Press.

Kliger, J., Singer, S., Hoffman, F., & O'Neil, E. (2012). Spreading a medication administration

intervention organizationwide in six hospitals. *Joint Commission Journal on Quality and Patient Safety, 38*(2), 51-60.

Laxmisan, A., Hakimzada, F., Sayan, O. R., Green, R. A., Zhang, J., & Patel, V. L. (2007). The multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. *International Journal of Medical Informatics, 76*(11-12), 801-811.

Li, S. Y. W., Blandford, A., Cairns, P., & Young, R. M. (2008). The effect of interruptions on postcompletion and other procedural errors: An account based on the activation-based goal memory model. *Journal of Experimental Psychology-Applied, 14*(4), 314-328.

Li, S. Y. W., Magrabi, F., & Coiera, E. (2012). A systematic review of the psychological literature on interruption and its patient safety implications. *Journal of the American Medical Informatics Association, 19*(1), 6-12.

Liu, D., Grundgeiger, T., Sanderson, P. M., Jenkins, S., & Leane, T. (2009). Interruptions and blood transfusion checks: Lessons from the simulated operating room. *Anesthesia and Analgesia, 108*, 219-222.

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2009). *The multitasking myth: Handling complexity in real-world operations*. UK: Ashgate: Aldershot.

Magrabi, F., Li, S. Y. W., Day, R., & Coiera, E. (2010). Errors and electronic prescribing: A controlled laboratory study to examine task complexity and interruption effects. *Journal of the American Medical Informatics Association, 17*(5), 575-583.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology, 14*, 127-144.

Merry, A., Weller, J., Robinson, B., Warman, G., Davies, E., Shaw, J., . . . Wilson, L. (2008). A simulation design for research evaluating safety innovations in anaesthesia. *Anaesthesia, 63*(12), 1349-1357.

Monk, C., Trafton, J. G., & Boehm-Davis, D. (2008). The effect of interruption duration and

demand on resuming suspended goals. *Journal of Experimental Psychology-Applied,
14*(4), 209-213.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a
nuclear power plant than meets the eye. *Human Factors, 42*(1), 36-55.

Neisser, U. (1982). *Memory: What are the important questions.* Paper presented at the
Proceedings of the International Conference on Practical Aspects of Memory, Cardiff,
Wales.

Pape, T. M. (2003). Applying airline safety practices to medication administration. *Medsurg
Nursing: Official Journal of the Academy of Medical-Surgical Nurses, 12*(2), 77-93.

Prakash, V., Koczmara, C., Savage, P., Trip, K., Stewart, J., McCurdie, T., . . . Trbovich, P.
(2014). Mitigating errors caused by interruptions during medication verification and
administration: interventions in a simulated ambulatory chemotherapy setting. *BMJ Qual
Saf*. doi: 10.1136/bmjqs-2013-002484

Raban, M. Z., & Westbrook, J. I. (2014). Are interventions to reduce interruptions and errors
during medication administration effective?: a systematic review. *BMJ Quality & Safety,
23*(5), 414-421.

Rivera, A. J. (2014). A socio-technical systems approach to studying interruptions: Understanding
the interrupter's perspective. *Applied Ergonomics, 45*(3), 747-756.

Rivera, A. J., & Karsh, B.-T. (2010). Interruptions and distractions in healthcare: Review and
reappraisal. *Quality & Safety in Health Care, 19*(4), 304-312

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking
continuum: From concurrent performance to task switching, interruption, and resumption
Proceedings of CHI 2009 (pp. 1819-1828). Boston, MA.

Sasangohar, F., Donmez, B., Trbovich, P., & Easty, A. C. (2012). Not all interruptions are created
equal: Positive interruptions in healthcare. *Proceedings of the Human Factors and*

Ergonomics Society Annual Meeting, 56(1), 824-828.

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Tomietto, M., Sartor, A., Mazzocoli, E., & Palese, A. (2012). Paradoxical effects of a hospital-based, multi-intervention programme aimed at reducing medication round interruptions. *Journal of Nursing Management, 20*(3), 335-343.

Trafton, J. G., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58*(5), 583-603.

Trafton, J. G., & Monk, C. (2007). Task interruptions. *Reviews of Human Factors and Ergonomics, 3*, 111.

Trbovich, P., Griffin, M. C., White, R. E., Bourrier, V., Dhaliwal, D., & Easty, A. C. (2013). The effects of interruptions on oncologists' patient assessment and medication ordering practices. *Journal of Healthcare Engineering, 4*(1), 127-144.

Trbovich, P., Howell, D., Easty, A. C., Ferguson-Pare, M., Koczmara, C., Savage, P., . . . Wiernikowski, J. (2010). Mitigation of interruption effects on delivery of routine high-risk medical procedures through design and implementation of effective interventions.: Canadian Patient Safety Institute.

Trbovich, P., Prakash, V., Stewart, J., Trip, K., & Savage, P. (2010). Interruptions during the delivery of high-risk medications. *Journal of Nursing Administration, 40*(5), 211-218.

Walter, S. R., Li, L., Dunsmuir, W. T. M., & Westbrook, J. I. (2013). Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours. *BMJ Quality & Safety*.

Weigl, M., Müller, A., Zupanc, A., Glaser, J., & Angerer, P. (2011). Hospital doctors' workflow interruptions and activities: an observation study. *BMJ Quality & Safety, 20*, 491-497.

Westbrook, J. I., Woods, A., Rob, M. I., Dunsmuir, W. T. M., & Day, R. O. (2010). Association

of interruptions with an increased risk and severity of medication administration errors. *Archives of Internal Medicine, 170*(8), 683-690.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science, 11*(4), 128-133.

Woods, D. D. (1985). The observation problem in psychology *Westinghouse Research and Development Center Technical Report*. Pittsburgh, PA.

Woods, D. D. (2003). Discovering how distributed cognitive systems work. . In E. Hollnagel (Ed.), *Handbook of cognitive task design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Woods, D. D., & Christoffersen, K. (2002). Balancing practice-centered research and design. In M. McNeese & M. A. Vidulich (Eds.), *Cognitive Systems Engineering in Military Aviation Domains.* (pp. 121-136). Wright-Patterson AFB, OH: Human Systems Information Analysis Center.

Acknowledgements

Author biographies

Penelope Sanderson obtained her PhD in engineering psychology from University of Toronto in 1985. She is Professor of Cognitive Engineering and Human Factors at The University of Queensland, with appointments in Psychology, ITEE, and Medicine.

Tobias Grundgeiger completed his PhD in psychology at The University of Queensland in 2011. He is currently working as post-doctoral researcher and senior lecturer in the Institute Human-Computer-Media at Julius-Maximilians-Universität Würzburg, Germany.

Table caption

Table 1

Fidelity, control, and potential generalizability of example papers.

Figure caption

Figure 1

The 12 highlighted studies placed within a fidelity/control/potential generalizability space.

Locations are approximate. Colors/shading indicate the form of investigation of each study.

Additional nodes whose heads are linked represent approximate locations of the phases of the

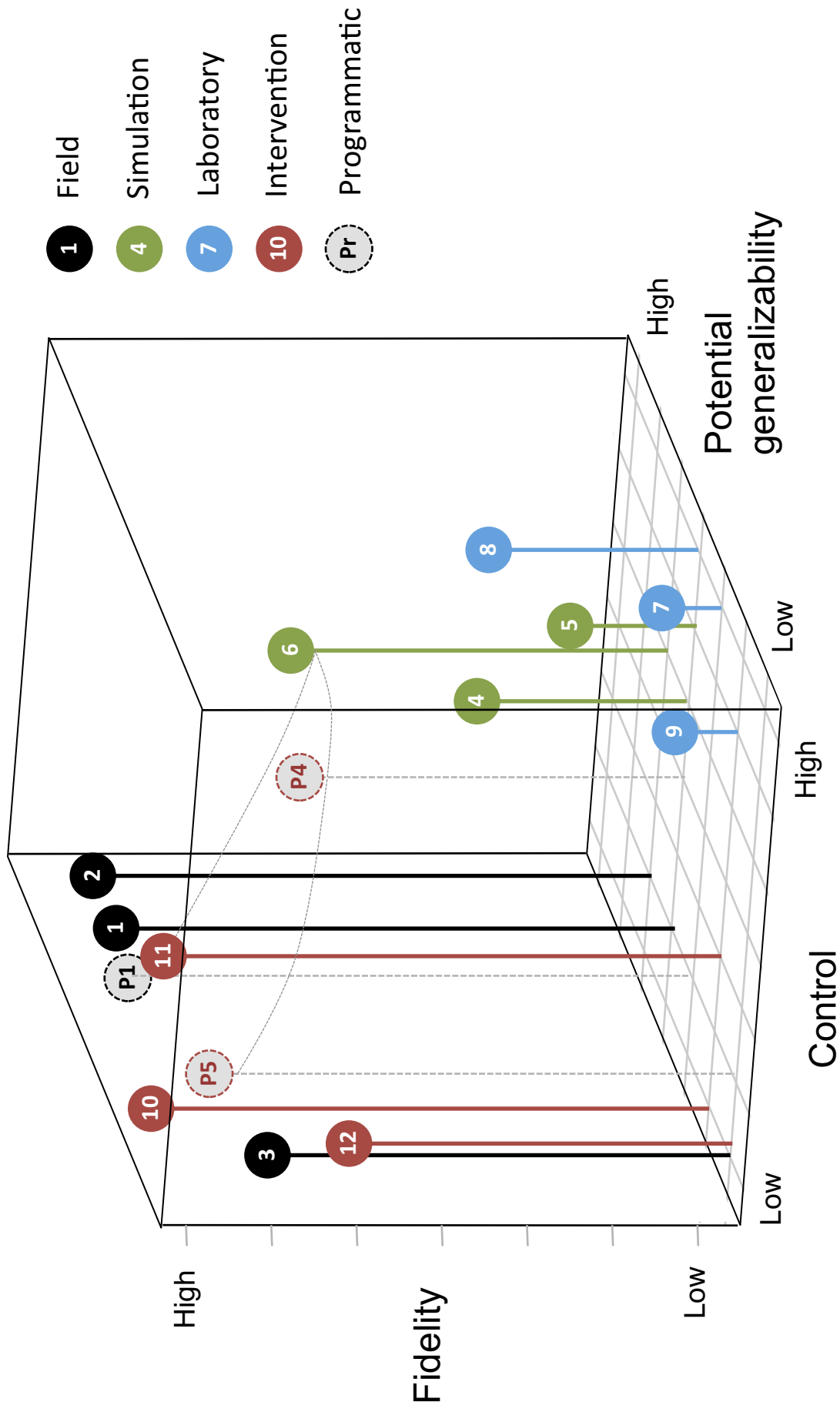Trbovich et al. (2010) program of research.

Table 1

Fidelity, control, and potential generalizability of selected papers.

Authors (year): Main research question	Fidelity		Control		Potential generalizability	
	Lo Hi		Lo Hi		Lo Hi	
Field						
1. Westbrook et al. (2010): Multi-site observational study seeking the association of interruptions and procedural and clinical errors during medication preparation and administration.	●	+ Nurses working in their habitual work environment.	●	+ Constraints on sampling. - No prospective control exercised to enable experimental contrast.	●	+ Comparison between hospitals. - Limited to interruptions during medication administration task. - No further refinement of task properties that could indicate generalization to other tasks.
2. Grundgeiger et al. (2010): Theoretically guided study using an eye tracker on the resumption of interrupted tasks and interruption management in an ICU.	●	+ Nurses working in their habitual work environment.	●	+ Constraints on sampling (i.e. patient condition, time of day), - Post-hoc experimental contrast	●	- Single ICU and only morning hours + Use of established theory (memory for goals) and refinement of task properties that influence task resumption. - Need theory of nurse management of interruptions
3. Rivera (2014): Qualitative study in an ICU to investigate nurses' decision to interrupt other nurses (observation and interviews).	●	+ Nurses working in their habitual work environment. - Interviews result in abstract situations.	●	- Single observer, interviews with focus on experience of nurses*	●	- Epistemological constraint from use of ethnography: "transferrability" not inherent to study but must be determined by reader alone.*

Simulation

Study						
4. Feuerbacher et al. (2012): Test whether operating room distractions and interruptions, [ORDIs] (present vs. absent) induce errors in a simulated procedure performed by novice surgeons.	●	+ Novice surgeons. + Scripted scenarios based on observations with specified ORDI points, and surgeons as participants. - ORDIs partly initiated by observer (who is not part of the scenario).	●	+ Experimental contrast. - ORDIs manipulation within participants (analysis of potential carryover effects not reported).	●	- Observer does most ORDIs and no further refinement of ORDI properties to clarify properties that have greater or smaller effect as ORDIs.
5. Magrabi et al. (2010): Test whether interruptions (present vs. absent) and task complexity (low vs. high) affect error rates when clinicians prescribe medication using a computerized provider order entry system in a simulation.	●	+ Doctors. - Part-task simulation of medication prescribing task with constrained behavior for participants (had to accept interruption) and initiation of interruption by experimenter (who is not part of the scenario).	●	+ Experimental contrast and detailed description and tight control of tasks. - Interruption manipulation within-participant (possible order effects).	●	- Experimenter causes interruptions and clinician has no choice about interruption management. + Use of established theory (memory for goals) and refinement of task properties that influence task resumption.
6. Prakash et al. (2013): Test whether interruptions make nurses less likely to detect planted errors during medication verification and more likely to commit errors during medication administration compared with no interruptions (pre-intervention)	●	+ Nurses. + Full-scale simulation of chemotherapy administration in an oncology ward. - Probably more errors planted than normally encountered in equivalent time on ward.	●	+ Standardisation of timing, actor behavior, nature of interruptions, and planted errors. + Conditions tested between-participants so no carryover effects.	●	+ Clinician has discretion over how interruptions handled. + Tasks, errors, and interruptions selected reflect prior field research. - Generalisability to non-oncology wards not directly addressed.

Laboratory

7. Brumby et al. (2013): Investigate the effect of resumption error costs and long task resumption times (i.e. resumption lags) on resumption errors using a donut-making microworld task.	● (left)	- Students. - Microworld task with steps and subtask that has not been specified in relation to a field situation. - Forced acceptance of interruptions.	● (right)	+ Experimental contrast and tight control of tasks.	● (left)	+ Use of established theory. - No description for which task the microworld and situation are representative - Participant cannot control interruption management or task resumption point.
8. Bogunovich & Salvucci (2011): Investigate the management of deferrable interruptions with small vs. large time constraints using a ringing-phone scenario.	● (right)	- Students. + Discretionary interruption management to some extent possible. + Familiar everyday task for participant.	● (right)	+ Experimental contrast and detailed description and tight control of tasks.	● (right)	+ Less constrained behavioral options for managing interruptions.
9. Cao & Liu (2013): Test whether auditory monitoring and/or a memorization task affect ability to perform a diagnostic decision making task, where diagnosis is classification into one of eight states based on three properties.	● (left)	- Students. - Forced acceptance of single- or dual-task condition. - Abstract diagnosis task plus auditory monitoring and memorization tasks with only superficial similarity to healthcare tasks.	● (right)	+ Experimental contrast and detailed description and tight control of tasks. - Multitasking manipulation performed within-participants (possible order effects).	● (left)	+ Use of established theory: dual task interference and automatic vs. controlled processing. - Situations for probable generalization of theory not systematically analysed or specified in detail.

Intervention

10. Tomietto et al. (2010): Pre-post, multi-intervention program to reduce interruption frequency during medication	● (right)	+ Nurses working in their habitual work environment.	● (left)	- Multiple interventions and pre-post design.	● (left)	- No comparison between units. - No further explanation of how interventions affect interruption rate.

rounds in seven surgical units.						
11. Kliger et al. (2012): Pre-post-post quality improvement intervention (among other the reduction of distractions and interruptions) to improve medication administration accuracy in six hospitals.	●	+ Nurses working in their habitual work environment.	●	- Multiple interventions and pre-post-post design.	●	- No comparison between hospitals. - No further explanation of how interventions affect interruption rate and what caused higher medication administration accuracy.
12. Pape (2003): Quasi-experimental study comparing the effectiveness of two interventions in reducing distractions and interruptions during medication administration.	●	+ Nurses working in their habitual work environment. - Research control of work situation and tools.	●	- Comparison of baseline and interventions in same unit (possible diffusion of treatment). - Possible order effects.	●	- No further explanation of how interventions affect interruption rate. - No basis for further generalization.

[+]Numbers next to paper author names refer to entry numbers in Figure 1.

*A qualitative approach usually eschews the exertion of formal experimental control, although sampling of respondents and roles may be systematic.