# Investigations Of Device Use In Radiology Departments

Rory Stuart Clark
[redacted]

Submitted to Swansea University in partial fulfilment
of the requirements for the Degree of Doctor of Philosophy

Swansea University
Prifysgol Abertawe

Department of Computer Science
Swansea University

7th November 2025

# Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed Rory Clark

Date 7th November 2025

# Statement 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed Rory Clark

Date 7th November 2025

# Statement 2

I hereby give consent for my thesis, if accepted, to be available for electronic sharing

Signed Rory Clark

Date 7th November 2025

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed Rory Clark

Date 7th November 2025

*And In The End The Love You Take Is Equal To The Love You've Made*
*(Lennon-McCartney)*

# Abstract

Radiologists are specialist doctors who examine visual medical information such as x-rays and CT scans to produce a diagnostic opinion for the treatment of a patient. Their work is extensively reliant on modern digital systems to generate scans, transfer and manipulate them for analysis, and communicate the findings to the relevant clinician. This means that radiology is often at the forefront of medical digitisation and implementation. There is, however, still a lack of holistic examination of precisely how – from an HCI standpoint - the radiologist interacts with the tools and devices at their disposal when constructing diagnostic reports and communicating them to others. This presents difficulties for the design, evaluation and implementation of new and novel capabilities.

This thesis examines the space in which a radiologist works from an ethnomethodological standpoint. We conduct studies in the form of ethnomethodologically informed ethnographies that examined the radiologist's interactions with digital and physical systems, allowing the reconstruction, analysis and critique of the design of the devices in-use from a holistic standpoint, with particular emphasis on the interoperability and usability from an outsider's perspective. These ethnographies utilise interactive and non-invasive observational techniques, allowing observers to gain an unbiased perspective on the interactional process without being ignorant to the medical and systemic influences in place. Supplementary to our ethnographies, we conducted a pair of focus groups with clinicians before and after examining the user's perspective on the issues they face and how they believe they can be addressed, thus allowing us to situate ourselves in the context and environment that we would be investigating and gather feedback on our findings and conclusions throughout.

This work provides a much-needed reconstruction and presentation of the ways in

which radiologists interact with and perceive digital devices when constructing and communicating their diagnostic opinion. We contribute an ethnomethodological analysis of the reporting process to allow for digital designers and clinical leaders to properly account for the real-world aspects of interaction that affect design and implementation of devices. Through collaborative and participant-driven methods, we also contribute codified and contextualised end-user clinical perspectives on how existing devices can be improved upon and how the design process can incorporate their domain expertise. Finally, we offer a reflective methodological contribution through exploration and analysis of the strategies when conducting research in a healthcare space as an Early Career Researcher in HCI, allowing researchers to improve the quality of future research in this domain.

# Acknowledgements

I'd like to begin by thanking Tom, Martin and Matt for their constant academic and emotional support throughout. I'd also like to thank Phillip, Tom, Victoria, Tracey and Caroline at the NIAW and Chris at the NDR for their endless enthusiasm and willingness to help in any and all capacity that was asked of them, and finally Sherryl and Ollie for their hard work in the CDT. This wouldn't be possible without them.

To Fergus, Matty, Jason, Beth, Anna, Emily, Saskia, Tulsi, Kyle, Luke, Beatriz, Anastasiya and everyone else I have met on this amazing academic journey, I'd like to extend a heartfelt thanks for their ability to always make me laugh, and to remind me why this is all worth doing.

To Chris, Dan, Owain, Ryan and Matt for their musical contributions to my life. Rock over London, Rock on Chicago. Also to Harry, Fred, Tom, Iago and Leon for the welcome distraction that is village cricket. I hope you never miss a straight one.

To Mum, Dad, Matt, Duncan, Lexi, Finn, Beth, Gosia, Kevin, Kerina, Steffan, Laura, Oliver and Ada, for the endless love and support, the importance of which cannot be measured.

Finally to my wife Lowri and, of course, to Mabel and Txula. For everything.

# Contents

# Preface

In 2019, I was offered the chance to make the move from the Humanities to the Sciences as part of Swansea University's Centre For Doctoral Training, funded by the EPSRC. This move required extensive training in the computational sciences, but presented a challenge and the opportunity of a lifetime. After completing my MSc, I was partnered with the National Imaging Academy Wales on behalf of the National Data Resource to investigate how their radiologists interact with the digital tools and devices at their disposal, a project I believed I could throw myself into wholeheartedly and genuinely *make a difference*. I am still eternally grateful for the opportunities afforded to me by the CDT, the NDR and the NIAW.

In this thesis, some stylistic choices have been made. This is partially owed to my academic background, and partially to the style guide to which I owe so much of my own writing techniques, *Style* by F.L. Lucas [187]. The majority of the research presented throughout is written in the first person, but using collective pronouns. This is for the most part a stylistic choice, and I wish to impress upon the reader that this thesis is solely the work of the author, unless otherwise specified. It is also, however, for comfort and personal reasons. I also wish to remind the reader that I have prioritised a naturalistic writing style over an overly academic one, as this enables me to better explain processes, motivations, and concepts.

# Supporting Materials

The works below include peer reviewed publications as well as abstracts, non-archival posters and presentations that contribute to the work found within the thesis. In all instances, the author of this thesis is first author and was responsible for the design and write-up of the research.

**Conference Papers**

- **Rory Clark**, Tom Owen, Matt Jones, Martin Porcheron, Phillip Wardle, Tom Micic, Bethany Delahaye: That Would Have Been Bad; How Radiologists Interact With Voice User Interfaces When Authoring Reports - *ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '25*), Bergen, Norway

- **Rory Clark**, Tom Owen, Matt Jones, Martin Porcheron, Phillip Wardle: It Works Better When I Do That; Interaction and Communication in Radiology Departments - *ACM Conference on Design Of Communication* (*SIGDOC '23*), FL, USA

**Workshop Papers**

- **Rory Clark**: Utilising Ethnographic Methodologies to Ensure Sustained Adoption of Digital Systems In Healthcare Environments - In *CHI'23 Workshop on Intelligent Data-Driven Health Interfaces (IDDHI '23).*

**Poster Presentations**

- **Rory Clark**, Martin Porcheron, Matt Jones, Phillip Wardle, Victoria Whitchurch: Perspectives On Machine Learning and Artificial Intelligence from Trainee Radiologists - *European Congress Of Radiology*, (*ECR '22*), Vienna, Austria

- **Rory Clark**, Martin Porcheron, Matt Jones: Ethnographic Methods For Studying Radiology - *NHS Cwm Taf Morgannwyg Research Conference*, 2022, Cardiff, UK

- **Rory Clark**, Tom Owen, Matt Jones, Martin Porcheron: Observing Radiologists in the NHS - *Festival Of Ideas Conference*, 2023, Swansea, UK

**Oral Presentations**

- **Rory Clark**, Tom Owen, Matt Jones, Martin Porcheron, Phillip Wardle: Communication in Radiology - *NHS Cwm Taf Morgannwyg Research Conference*, 2023, Cardiff, UK

**Doctoral Consortiums**

- **Rory Clark**: Ethnographic Investigations Of Device Use In Radiology Departments - *INTERACT 2023: 19th IFIP TC13 International Conference*, 2023, York, UK

# List of Tables

# List of Figures

# Chapter 1

# Introduction

According to the most recent Diagnostic Imaging Dataset Statistic Release, between November 2023 and October 2024 there were 48.5 million imaging tests in NHS England alone [112]. Almost every patient who requires diagnostics is run through the radiology department of the hospital, responsible for these imaging tests. The role is vital to the proper running of the hospital, and the average NHS radiologist is severely overworked – the 2023 census estimated that the workforce would need to increase by 10% overnight in order to properly address the volume of backlogged reports [225]. In Wales specifically, the number of radiologists per 100,000 people is 31% lower than the rest of the UK, with every clinical director across the country claiming they were unable to manage the reporting requirements of their health board [240].

In an effort to the mitigate this, the growth of digitisation in the hospital has had a significant impact on the duties and conduct of the radiologists, with most scans now being taken, delivered and analysed digitally. Reports are made on computers, and are mostly communicated to the referring clinician through cloud-based servers [298]. The radiologist is reliant on their personal computer, their internet connection, and their faith that these digital systems are accurate. Anything ranging from a major software crash to a single mis-transcribed word on an x-ray report can have serious ramifications to the timely and effective treatment of a patient [65]. This marks the radiologist's office as a *safety critical environment* [167], where all aspects of hardware, software and peripherals come together to create a sensitive environment where any minor failure can disrupt and disturb the radiologist at work, creating an error that could have serious consequences

[200, 201].

All of these medical devices have, of course, been individually deemed safe enough for the practitioner to place their trust in their day-to-day operation [131]. However, for such an important aspect of the hospital workflow, there is a lack of holistic examination of how these devices work once implemented in the real world as part of a singular diagnostic process. A mis-transcribed word may not simply be a fault of the transcription software, but instead the result of a concatenation of factors that result in a mistake [16]. This work examines the behaviour and practices of the radiologist at-work with a wide lens from a *Work As Imagined* perspective [151] (referring to the difference between practice in a perfect, laboratory environment and practice in a more real-world and pragmatic environment) to examine how they interact with the digital systems at their disposal in order to improve their design, and improve the satisfaction of them as the end-user. By exploring off-the-shelf devices that have been procured, approved for use and implemented in an actual work environment we can work to understand overt design flaws and interactive patterns, allowing for a more complete representation of the radiologist.

## 1.1 What Is A Radiologist?

*"A radiologist is like an F1 driver – you just want to be able to get into the cockpit, plug and play, and not have to worry about if anything is going to go wrong"* [1]

This quote from a consultant radiologist does well at contextualising the role of the diagnostic clinician when they are in the throes of analysing a medical study. The art of examining medical studies is one that requires years of perfecting, and is the culmination of decades of training and skill. It is extremely intensive work that requires the highest levels of attention, and mistakes can be deadly. Much like driving a vehicle at 200 miles per hour, the sheer cognitive load of the job means there is very little scope for distractions unless necessary. Here, we will briefly explain what a radiologist is, what they do, and the state of being a radiologist in the modern NHS.

The term "radiology" refers to a discipline of medicine that utilises medical imaging to diagnose patients. There are several modalities of radiology that rely on different

[1] Taken from a comment made by a participant in a study reported in Chapter 7

methods of scanning these patients - X-Ray, Ionised (density based, such as fluoroscopes and Computer Tomography or "CT"), Nuclear (Gamma camera and Position Emission Tomography), Ultrasound and Magnetic [221]. Within the radiology department, there are many positions that each have a different role to play in the patient's journey: Radiology nurses, medical physicists and health care support workers are all vital to the functions of the "imaging and reporting factory", but they will not hold specific relevance to this project. Instead, we will be focusing on the radiologist and the radiographer.

The radiologist is a specialist doctor, meaning they have qualified to be a medical practitioner and then spend roughly 5 years training in radiology [226]. There are three specialities within radiology: Diagnostic, Interventional and Therapeutic. Diagnostic radiologists utilise studies like X-Rays or CT scans to provide a clinical opinion of diagnosis that a referring clinician can use to administer treatment. As such, they often to not have direct interactions with the patient involved, and will often administer treatment guidance without having seen the patient at all. Interventional radiologists perform minimally invasive procedures on patients with the assistance of imaging devices. This includes removing clots, draining abscesses and performing biopsies. The therapeutic speciality involves administering radiation therapy to patients to treat cancers and associated diseases. This thesis will focus specifically on diagnostic radiologists.

The pathway of the "radiological factory" refers to the flow of information and guidance through the radiology department. It begins with a request for a referral from a clinician who wishes to investigate a patient, which is then either approved or declined by the radiologist in question; since medical imaging is time consuming, costly, and a serious procedure for the patient involved, the radiologist is the deciding factor in whether or not a scan is necessary, and will only approve a request if appropriate. This is known as the "vetting justification protocol". If approved, the patient is then taken for the relevant scan, which is then added to the Picture Archive and Communication System (PACS) and Radiology Information System (RIS) along with the reason for the request and any clinical notes on the patient. It will then be selected and reported on by the radiologist or radiographer, who will examine the study with a specific focus in mind (based on the request), and will finish their report with any incidental findings that may be important to the referring clinician. This report will then be uploaded back to PACS for the clinician

to read and treat the patient accordingly. The process then begins again, with a reporter examining multiple reports in a session [221].

## 1.2 Problem Description and Research Questions

When examining the radiological factory above, it becomes clear that the radiologist's working space is a safety critical environment, and devices must be implemented and evaluated accordingly; a holistic understand of the entire environment (digital systems and physical hardware included) is required when examining the ways in which reports are constructed and communicated. Computers are required to receive, load, investigate images, to record the radiologist's medical opinion, to communicate with other members of the hospital, and every aspect of the duties of the diagnostic clinician [298]. It becomes equally clear the level of training, education and tacit knowledge that the radiologist possesses when analysing and constructing a diagnostic opinion. Many systems and devices have to work together in collaboration with both each other and the user to perform these tasks, and the interoperability of everything is imperative to proper and sustained satisfaction and high performance when diagnosing a medical study.

The devices in use in radiology departments are subjected to analysis, testing and safety metrics in both the design and procurement process [131]. There is, however, a distinct lack of understanding regarding how all of these devices and systems (physical and digital) come together *holistically* [7] - it is therefore impossible to expect designers of radiological equipment or practitioners themselves to create and maintain an effective safety critical environment without this comprehensive mapping and understanding underpinning design and implementation.

These presents the need for an accurate, up to date and usable understanding of how radiologists interact with the medical devices at their disposal in a holistic way, with attention paid to every aspect of the reporting process. This understanding has to be a true representation of what actually happens in an imperfect and working environment utilising both our own impartial investigations and the tacit knowledge of users to allow for designers, developers and implementers to understand the context and situation in which the radiologist is interacting with every device. These designers should be building

tools and devices with a *Work As Done* frame of mind, as opposed to a perfect *Work As Imagined* one [151]. Thus, our problem statement exists in these 2 dimensions:

1. *How Do Radiologists Interact with Devices In-The-Wild When Constructing Diagnostic Reports?*

2. *How Can the Perspectives of Radiologists Be Utilised to Meaningfully Impact the Design and Implementation of Devices In Safety Critical Environments?*

## 1.3 Research Contributions

In this thesis, we offer the following research contributions:

### 1.3.1 An ethnomethodological reconstruction of how an NHS radiologist constructs a report in-situ

Through a series of ethnographic methods, we provide the reader with a series of vignettes that come together to construct multidimensional representation of the behaviours, practices and interactional processes that go into authoring a report. This representation presents a realistic and accurate portrayal of the user's interactions with the digital and physical systems at their disposal, as well as a wider social picture of how these interactions are developed and influenced by members of the hospital's social organisation.

This presentation of interactions and behaviours holds value to several stakeholders. It allows the diagnostic practitioner to reflect on their own practices, prompting discussion of best practice and techniques; it allows the wider NHS organisation to ensure that their procurement, adoption and implementation guidance is appropriate and holds the most accurate and holistic picture in mind; it provides designers and developers with a proper, real-world picture of Work As Done in-the-wild, allowing for tools and devices to be built safer, and more appropriate for their use in realistic scenarios.

### 1.3.2 A clinical perspective on where elements of friction and failure exist when diagnosing and treating patients, and how digital and non-digital interventions should be designed and implemented to address them

By frequently involving the user in our studies, we can provide insight into the ways in which practitioners perceive the problems that they encounter on a day-to-day basis are created and solved. We adopt a "checking-in" philosophy of research, by continually utilising methods such as focus groups, interviews and collaborative analysis, ensuring that our findings and conclusions are reflected in the experiences of the groups that we are examining.

From this, we present several findings in this thesis. We demonstrate a case of defining a research avenue through participatory study whereby informed and experienced members of the clinical domain provide areas to investigate based on their own knowledge in Chapter 4 and 5, and how clinical practitioner's expertise can be utilised to frame findings by HCI researchers; in Chapters 4, 5 and 6 all design recommendations are shaped by presenting findings back to participants either through post-participatory interviews, collaborative analysis sessions or through design-based focus groups utilising the thesis data as a focal point. This results in findings and recommendations initially presented from an HCI perspective that have been given an additional layer of insight by those who are actually likely to see the results of the work.

As such, the contributions of this thesis go beyond the HCI researcher entering an interdisciplinary space and providing findings based on their observations; we contribute more nuanced, actionable and practical design recommendations by utilising the clinical practitioner's perspectives on our findings after the data gathering stage before we present them to the reader.

### 1.3.3 A series of design and pedagogical recommendations for those constructing, implementing and using devices in health-based safety critical environments

Throughout this thesis, we use information based upon our own observations and analysis alongside participant-led perspectives to provide the reader with design recommendations

that will allow for more appropriate and safer adoption and use of devices and peripherals in radiology settings. We demonstrate the ways in which current design is unsuitable for its purpose and leverage our *work as done* perspectives to offer ways in which we believe future devices in these spaces can be improved upon.

Alongside this, we provide a series of recommendations to those teaching in these environments; Chapters 5 and 6 reveal significant gaps in the user's knowledge of how these devices work, and best practices for how to interact with them – as such, we advocate for better and more effective training on how to interact with the devices at the practitioner's disposal based on our findings, as well as a more general call for better recognition of the importance that peripherals and devices play in the treatment of patients.

As such, we not only provide contributions in the form of theoretical findings in terms of what constitutes interaction in the radiology office or in the form of methodological findings by incorporating a clinical perspective on our results, but we also provide tangible, actionable instructions and guidance to those operating in these spaces.

## 1.4 Thesis Structure and Outline

Aside from Chapters 1-3 and 9-10, we have chosen to present this thesis chronologically in how the studies were planned and conducted, as this provides the thematic journey that we found ourselves on, and enables a cohesive narrative of the development of our research contributions. Chapter 2 outlines our research focus and methodologies, and Chapter 3 provides an overview of the existing research and the gap that our studies fill.

We begin at Chapter 4 with a participant-led approach that allows us to understand the space in which we are working, establishing a level of "vulgar competency" (appropriate knowledge of the context that we are studying [183]) that will be required for the following studies . This participant-led exercise establishes a series of faults and failures that participants experience, focusing on the lack of high-quality and well implemented digital systems that practitioners feel they need when diagnosing and treating patients.

We use these perspectives to narrow down a non-interactive and indifferent study of general practice in-the-wild in Chapter 5. This observational study investigates the real-world

practices that go into constructing and communicating a radiology report by actually visiting the department and taking notes from practitioners as they work. This not only constructs an accurate picture, but provides us with impartial data to contrast against the claimed points of failure in Chapter 4.

Of particular note in our observations was the unique ways in which practitioners interacted with the voice-based transcription software when authoring reports. This led us to conduct a dedicated, more controlled investigation into this practice, as it was the hinge point for all aspects of the radiological process – the report was the sole product of the practitioner, and the VUI was how they made it. Chapter 6 consists of a laboratory-based study using real-world data and systems to gain a better insight into the ways in which practitioners use these systems, alongside a collaborative analysis exercise with a selection of participants.

The output of this study provided us with some design insights on how to address potential areas of fault and failure identified in Chapter 4, and so as a book-ending exercise we conducted another participant-led study in Chapter 7 that used our main findings in the observational and laboratory study as the prompts and discussion aids, alongside design mockups that came out of the results of these studies as a focal point. The results of this study provide us with a small, circular design process that demonstrates how the perspectives of practitioners, presented back to them after several research iterations, can end up contrasting with what they thought they would want.

Chapters 8 and 9 bring together all of this research to re-address our research questions and contributions, allowing us to summarise the thesis in a way that presents the reader with the so what of the research found within.

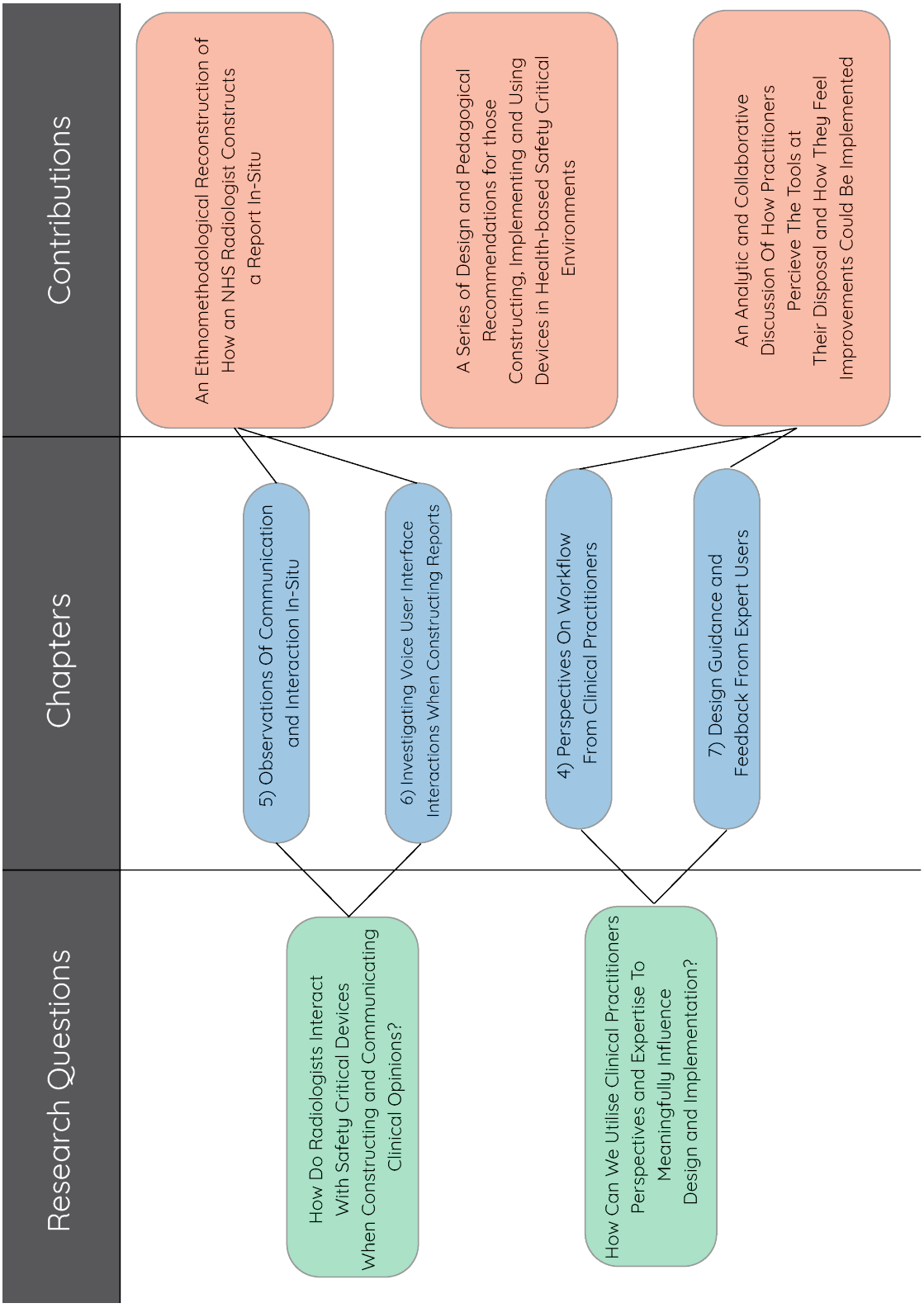


Figure 1.1: Colour coded outline of how our chapters align with RQs and contributions

# Chapter 2

# Related Work

## 2.1 Introduction

The aim of this chapter is to inform the reader of the existing work in this space and establish the research niche in which the work sits. As shown in the graphic below, this thesis sits in the intersections between Participatory and Collaborative Research; Work As Done vs Work As Imagined; and Diagnostic Medicine (specifically radiology). We will demonstrate here that work that addresses all of these at once does not currently exist, and we will provide an explanation for why this is necessary.

As each chapter contains a brief exploration of the motivation behind its content, this chapter will focus on core concepts, historic and narratively necessary case studies, and structural pieces of seminal research required to understand the work properly.

Through this chapter, we will establish that

1. Human factors is an important aspect of the safety critical environment, especially when assessing and evaluating the efficacy of devices, but can be overlooked from a holistic perspective

2. Radiology is a heavily digitised and technologically reliant discipline of medicine currently in the midst of rapid evolution, but its non-patient facing nature can exclude it from critical examination

3. In-the-wild and ethnographic examinations of device interactions has a rich history in HCI and health studies, but has not been applied to radiology in appropriate depth

As such, this thesis proposes to solve the problem of examining this technologically focused medical domain for human factors aspects of device interactions with ethnographic and ethnographic methods, effectively addressing the research niche.

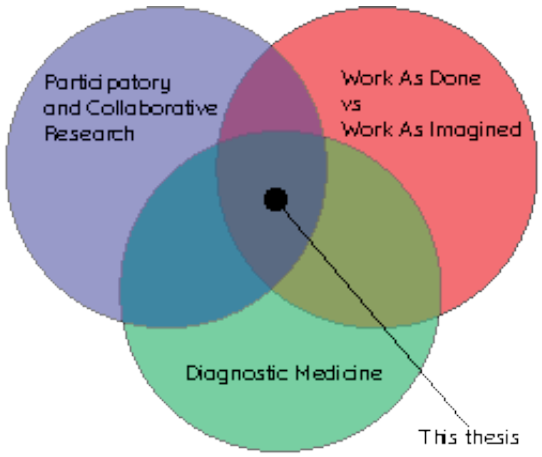


Figure 2.1: Visual Demonstration of Where This Work Sits In Related Literature

## 2.2 Work As Done vs Work As Imagined

*"Situated studies help the researcher to focus on the aspect of technology design and use that really matter to people, and that have an impact on practice."*
- Blandford et al, *Strategies for conducting situated studies of technology use in hospitals* (p. 491), 2015 [40]

Here, we will examine the current state of *Work As Imagined* by first examining HCI and Health in general before exploring more specific topics - the aim of this section is to provide the reader with an overview of what has (and hasn't) been done in this space before highlighting the lack of radiological focus.

### 2.2.1 HCI, Holistic Study and Healthcare

Human Computer Interaction and Healthcare have been closely related since the early days of HCI as a formal practice. Leveson's seminal inquiry into the major failures of the THERAC-25 radiotherapy machine revealed issues in the design of the user interface, demonstrating the need to understand interaction patterns and techniques when considering the implementation of computational devices in the medical space as early as the 1980s [179]. In their overview of Computer Supported Cooperative Work and Health based research from the 20th to the 21st century, Fitzpatrick and Ellingson identify Engestrom et al's 1989 paper *Computerized Medical Records, Production Pressure and Compartmentalization in the Work Activity of Health Centre Physicians* as the earliest example of a medically focused piece of research published in the journal CSCW - the work examines electronic patient records and their use in social organisation of the hospital [1] [118]. Fitzpatrick and Ellingson's review of CSCW and health is now over a decade old, and a more modern critique of the landscape from Andersen reveals a significant growth in the use of Artificial Intelligence in the domain, but a significant drop in implementations and examinations of AI in-the-wild [12]. Andersen opines that this gap is due to challenges in capturing the holistic view of evaluating the efficacy and role of modern technological systems in clinical workflow [12]; this factor has been acknowledged by the NHS as early as 2004, where Clarkson's report identified that, when designing context-specific tools and devices, there is a need to understand the healthcare environment in an all-encompassing manner before filtering down to individual interventions [82].

Today, there is an understanding that barriers to comfortable and efficient use of medical devices can appear over time, when workflow and practice have to accommodate for new technologies [202, 303, 304]. Modern work has further emphasised the need for holistic evaluation of interaction, organisation and digital interventions in medical environments [170, 110], with Kip et al's 2022 paper explaining that, in the current landscape, the variety of products available in the eHealth space make it possible to examine implementation in a too fine grain, not examining the wider implications of specific tool adoption [165], and Campbell's 2020 examination of telemedicine advocated for mixed-methods methodologies to capture a wider image of usability of digital devices [191]. Overall, however, there

[1] Demonstrating that, even after 35 years of dedicated study, the computational sciences are *still* debating the efficacy of EHRs.

is a lack of holistic views when examining HCI in medical situations - a literature review of the safety and usability of Clinical Information Systems showed only 2 publications from the pool of 76 that examined implementation with a wider lens [178].

### 2.2.2 Workplace Study

Workplace studies are a key genre of studying HCI and CSCW, as they allow for examination of how systems are interacted with in naturalistic settings, and can be undertaken with a variety of methodologies and approaches. In the 1990s, Hughes posed the 2 key arguments for moving study out of the lab and into the real world; there was a growing plausibility that systems fail due to insufficient attention to the 'social context' of work, and increasing awareness of the ubiquitous nature of networked and distributed computing and the problems it poses for collaborative and social activities [155]. In simpler terms, workplace studies come from a concern that technological systems are implemented without a knowledge of how ordinary people work [188]. This work is primarily ethnographic in nature, examining existing systems that have already been implemented [189], and has its earliest roots in Suchman's *Plans and Situated Actions*. Suchman conducted a series of studies where a XEROX photocopier was the subject of ethnographic analysis, already demonstrating that so-called "ordinary" aspects of everyday life deserved to be the topic of HCI scrutiny [291].

When regarding Safety Critical situations, workplace studies tend to emphasise *"On Land, At Sea, In The Air"* [201] referring to the high volume of control-room-based research that has been undertaken in situations such as energy power stations, cockpits, ship bridges and intensive care wards [200]. Research in this area tends to focus on areas with organisational and management capacity, like Air Traffic Control towers [36], Nuclear Power Control Stations [69], and the Engine Control Room of large ships [194], and can reveal significant ergonomic, layout and software design flaws that only become apparent when natural and "normal" patterns of behaviour are observed and analysed.

As such, a domain that often comes under scrutiny of a workplace study is the hospital, with a common area of focus being organisation and communication among and between groups. As early as the 1990s, observational studies of communication and coordination *in-situ* were being conducted to examine areas of failure in communications

between doctors and nurses - Coiera et al's work revealed communication techniques were based on personal preferences and a dislike for written forms of communication, leading to interruptions when preferences clashed [84]. Reddy et al similarly utilised fieldwork in the workplace to gain an understanding of the *rhythm* of an emergency ward, allowing them to map out the coordinated and unspoken patterns of behaviour that would have been otherwise inaccessible [250, 251]. Other work has helped reveal areas where information sharing breakdowns occur [2], the importance of annotation and physical documentation when keeping track of patient treatment [196] and identify error rates in patient handoff and outpatient workflow [272]. A particular study of note examining hospital orderlies and their role in organising and facilitating patient treatment and workflow highlighted the importance of workplace study for promoting and increasing the visibility of "invisible" work that didn't directly impact the treatment of patients [287]. Often, these studies can reveal that particular aspects of a job are learned intuitively, despite their importance to the role - Brassuer's study of sonographers in their treatment rooms demonstrated that, despite a high amount of their duties involving delicate and sensitive information having to be communicated to concerned or anxious patients, they received no education in how to speak to their patients [54].

These studies show that "*being there*" and examining examples of real work taking place in-the-wild is an important method of understanding and mapping patterns of organisation, communication and cooperation in the hospital. This lens has been focused on the radiologist before (albeit in limited fashion) in work by Briedis from an solely phenomenological perspective, unpacking the complex sociological praxis that goes into the spatial arrangement of the American radiologist's office and environment [57, 58]. Briedis' work here is reflective of our own goals, in studying the interactions and tacit behaviours of the radiologist at-work, but they come from the perspective of the psychological mental models at work when analysing and examining medical studies. As such, it is a useful examination for us in comparing our work in NHS hospitals to theirs in North American ones, but we offer a series workplace studies that examine the role that technology plays in facilitating and assisting interaction, having a focus that is on technique and design. It is, however, useful to see that the importance of examining the radiologist at-work in their natural environment is recognised; Briedis claims that between their own work and that of Alac [7], there has been close to no study of radiologist praxis

- this is clearly not entirely true, as the work of Johnson revealed patterns of communal work and distributed interaction with devices when conducting patient treatment in interventional radiology [161], and demonstrates the potential disconnect in awareness between sociological investigations and technological ones when studying radiologists. Work such as ours helps to bridge this gap and bring the two together.

### 2.2.3 Human Factors Engineering and Design

In their essay on ethics in safety critical design, Bowen claimed that is is *unethical* to develop and implement software for safety critical systems without maximising the level of expertise on a project - they claim that all design should be grounded in as much knowledge as can be reasonably ascertained [50]. As such, Human Factors engineering is a commonly utilised method of *complimenting* traditional mathematical formal methods as a way of understanding and improving safety critical design [134]. As part of a holistic approach to implementing and maintaining a safety critical system within a high pressure and dynamic environment, human factors is a key aspect of the mixed-methods approach to orienting teams and designers towards better usability and safety goals [163]. This means that, alongside formalised grounded methods of ensuring system safety, we have an understanding of the interactional and behavioural patterns of those using these systems, to allow for a multi-faceted approach for mitigating error and failure.

As expected, human factors engineering concepts has been often applied to medical environments. Analysis of clinical workflow and areas of failure are common for understanding points of friction, but work by Niazkhani claims that there is a distinct social and organic nature to collaborative medical work that can only be reckoned for when factoring in human behaviours into safety evaluation and assessments [222]. Similarly, Greenhalgh's work on Electronic Health Records and digital workflow demonstrates that, unless technology conforms and aligns to personal attitudes and routines, there will be significant barriers to adoption that will lead to abandonment from independent clinicians [132, 133]. The growth of machine learning and artificial intelligence in these spaces has also produced an increase in human factors and user centred design - Beede writes that, regardless of the quality and efficiency of a deep learning algorithm, the requirement for nurse and clinician education, infrastructure, financial and temporal resources and other supporting factors all play a significantly higher role in the quality of care that

the algorithm may be able to provide [32], and Bucina's work on AI in clinical decision support shows a split between over reliance and a complete lack of trust when presented with an artificially generated recommendation, with the problem lying in the lack of consideration regarding current working practices of practitioners [62], echoed in Cai's work highlighting the ways in which clinicians want to treat AI models like co-workers, an aspect not considered by designers [64].

Outlined here, the key benefit of human factors engineering to safety critical environments is *understanding behaviour*, as this allows for a more in-depth and nuanced understanding of how the medical ecosystem works. There are drawbacks to focusing solely on user centred aspects of design; Hohm identifies that, as good as UX is in highlighting the *eudamonia* of a system, it causes divergence in what some designers can consider "safe" [149], and the best ways to include user feedback in the medical device design and implementation loop have still not been properly codified and agreed upon [180]. Additionally, the barriers to implementing User Centred Design in health spaces can be systemic - Vincent outlines that, alongside having to understand and map collaborative work practices in diffuse environments, there can be hierarchical friction in explaining *why* user centred design should be accommodated for, and providing adequate justification for the adoption of human factors is a significant barrier to their sustained implementation in the design of medical devices [308]. There are ways to address this - a *lessons learned* case study from Power et al of the development of an NHS safety thermometer demonstrated that, despite systemic issues and infrastructure problems when gathering user perspectives and meaningful ethnographic data, it is feasible to utilise user centred design methodologies to produce actionable changes to safety critical medical devices by adapting techniques to accommodate for regional variation and sample sizes [238].

### 2.2.4 Error In Safety Critical and Medical Environments

Dekker describes human error as '*A Window To A Problem*', referring to the their perspectives on what human error represents when investigating a failure: it is not a cause, an effect or a symptom, nor is it the conclusion - human error merely shows the starting point of a failure [96]. Booth identified the problems in determining error in safety critical situations as the clash between understanding of the failure and the broadness of error theory [47]. Whilst this identification was based around error theory of the 1990s, it is clear

that the difficulties are still present in modern healthcare - the play-off between requiring deeply complex and nuanced safety critical systems and a need to have them robust enough to accommodate human factors is one many medical environments struggle with [197]. Blandford claims there are 6 key reasons for an interactional failure in a medical environments: a lack of human factors guidelines in the procurement process; a focus on patient care over device design; minimal questioning of device design; a pervasive blame culture in investigating failures; the lack of high quality data for design and testing; and a lack of ownership in the problem [41].

As mentioned, a common issue with addressing and mitigating errors in medical situations is the nature of identifying errors and their root cause. Detection of errors in real world situations is complex in itself [150], and in many cases, *near misses* or errors that do not result in harm are not reported, meaning potential design and interaction flaws go unnoticed until they compound to a serious sentinel event [24, 17]. An investigative study with nurses in the USA found informal rules for determining if an event can be classified as an error that requires reporting: if it is not my fault, if everyone knows about it, if it can be quickly fixed, if the patient's needs outweigh the administrative burden of reporting it, if it was simply a clerical mistake or if it was something that prevented an even larger problem, it should not be reported as an error [21]. These informal rules demonstrate the difficulty in identifying and addressing interactional or design problems that result in errors based purely on recorded data - Boll claims that medical design is based upon a *lessons learned* approach, meaning a holistic understanding of how error gets introduced into a clinical workflow is still something lacking in design spaces [45].

### 2.2.5 Work As Done vs Work As Imagined

Having highlighted the validity and utility of the workplace study in revealing organisational patterns in the hospital, we can examine its importance in uncovering Work as Done vs Work as Imagined in terms of device interaction in-situ. *Work as Done vs Work as Imagined* as a topic of study in its current and modern recognisable form began in 2004 as a way to pay attention to and highlight *'informal practices that have become formalised'* [53]. In examining the difference between 'procedure' and 'practice', the goal in Work As Done (WAD) work in HCI and CSCW is to illustrate the ways in which behaviour and interaction diverges from the designer and manufactures intended or believed methods

of use in-situ, with the hope that this will result in reflection and future iterations of devices and systems that are more user friendly and resilient to error and failure [151]. Much like the more general implementation of workplace study, WAD work has been found to be effective in understanding ways in which real-world behaviour deviates from what is written in procedure in Safety Critical environments such as Air Traffic Control [316, 70], Pharmacy dispensers [18], Nuclear Power Stations [68], and firefighters whilst out on jobs [71]. An example of ethnographic WAD fieldwork study producing previously unseen yet important insights is Borys et al's work in construction sites - their research revealed that the increase and ubiquity of paperwork required to undertake and complete a construction job meant that managers and foremen on-site had resulted in a cultural separation and an *illusion of high levels safety* of the workers, leading to an increased level of potentially harmful behaviours appearing on site that were unreported [49].

WAD has particular benefits in medical contexts, where it can reveal poor alignment between managers and practitioners resulting in a lack of high quality implementation [235, 236], and examples of misuse that can result in patient harm. In their scoping study, Blandford et al found "workarounds" (i.e. ways of making devices do things they were not designed to do, or circumnavigate safety procudres) in variety of contexts including patient information barcodes being improperly applied to furniture, generic codes being input into glucometers, and syringes being used to administer milk to infants [42]. A common example in literature as a result of the CHI-MED project is infusion pumps; shadowing and observational work of nurses using infusion pumps has revealed that "unremarkable" errors, or small seemingly low impact mistakes that practitioners did not report, can often compound into serious issues and cause harm to patients [122], and analysis of infusion pumps over time revealed that at least 44% are interrupted at least once over the course of administering medicine to patients [3].

The difficulty comes in **addressing** WAD in medical settings. Studies such as Taylor, Damoiseaux-Volman, Kislove and Rayner help to highlight gaps in WAD, but do not offer guidance, instead simply highlighting that *there are gaps* and more work should be done in these spaces [297, 91, 247, 236]. Vincent describes the challenges in closing these gaps as a lack of scenario-based design examples, the difficulty in mapping and covering breadth of device use, and managing the contradictory evidence that comes

from user-centred design methodologies [309]. A common method of addressing poor implementation and design failures is examining other examples of safety critical environments - examples of research that extracts techniques from the aviation industry and suggests their implementation in medicine are easy to find [253, 232, 299, 274, 312], but critical reflections on this comparison have become more widespread in recent times, with the common conclusion being that whilst there are some parallels, looking to other safety critical environments is not appropriate for a nuanced and highly specific domain like healthcare [106, 104, 156, 162, 248]. As such, there is a need to come up with **solutions** to areas of friction when examining WAD, and assess their efficacy in-situ.

### 2.2.6 Research Gap

What we demonstrate here is that the concept of "*being there*" and studying interaction in-situ has a long history with HCI and its effectiveness has been shown in a multitude of environments. *Work As Done* study has a specific efficacy when applied to medical contexts due to the dynamic nature of providing care in emergency situations and the deviations from textbook-style practice.

We have, however, highlighted 2 areas of attention - first, that radiology has suffered from a distinct lack of individual examination from an HCI perspective despite its reliance on technological systems; Briedis' claim that only their own and Alac's work [57, 58, 7] highlights the ways that this non-patient facing discipline has been overlooked. Second, that whilst WAD study has been shown to be effective in demonstrating areas of failure and friction, it often fails to provide effective guidance to address the problems it presents.

## 2.3 Participatory and Collaborative Research

*"Situating a new technology within a narrative forces us to grapple with questions of ethics, values, social perspectives, causality, politics, psychology, and emotions"*
- Tanenbaum, *Design Fictional Interactions: Why HCI Should Care About Stories* (p. 22), 2014 [293]

This section examines the method of utilising participant data and perspectives as a key point of leverage when assessing practice and constructing guidance - we demonstrate

the need for end-user's viewpoints when conducting HCI research, and the utility of having clinicians as participant-researchers to utilise their tacit knowledge and high levels of experience.

### 2.3.1 Participatory and Collaborative Design in Healthcare Contexts

Hansen describes the distinguishing feature of participatory design is in the mechanisms used and their implications for democratisation in use and design of technology - it is not simply the output of the activity that holds value, but the ability to educate and promote engagement with communities [141]. In similar words, Kettley et al claim that by utilising a person-centred and collaborative approach to design, we can allow for a more *dynamic* and *holistic* examination and mapping of experiences with tools and devices [164]. These two descriptions provide an explanation for why participation from clinicians and patients is a growing method of examining implementation and design of medical *devices*, and medical *environments*.

Primarily, participatory and collaborative methodologies in these domains have been used as a method of promoting engagement or prompting further discussion in a subject area [38, 231]; studies have demonstrated factors such as including parental and guardian figures can increase knowledge around medical interventions for children [267], increasing the visibility of clinicians as part of the design process [25], promoting doctor's voices and opinions over the procurement and implementation of new medical devices [102]. Of particular note, Wong et al found that utilising clinical perspectives in the *post-implementation* stages of device use was valuable in evaluating the standards of e-health technologies [317, 318].

There has also been considerable interest in the *outputs* of participatory design in healthcare contexts, from designing mobile apps for patients [207, 275] to using experience and perspectives to identify areas of failure with medical devices [223], designing devices for irregular environments such as care homes [137] and creating materials for education and information on personal health [93]. These studies help to demonstrate that, alongside the communal and potentially unquantifiable personal impacts of utilising participatory and collaborative design demonstrated above, there can be actionable and measurable benefits to the method in tangible formats.

There are, however, significant caveats to utilising participatory and collaborative design in this domain. Experience-based Co Design is a method frequently used by the NHS to get patients and clinicians to collaborate on projects, with the aim to bring experiences of patients receiving care and expertise of practitioners together [101]. Whilst it can produce useful outcomes, the key critique of the method is that it creates significant friction between the two groups due to the uneven power dynamics [51], with patients seen by medical participations as not understanding clinical perspectives and clinicians seen as condescending by patient participants - this is also found in co-design activities between practitioners and researchers, with the power dynamics reversed where highly qualified domain experts are seen merely as data gatherers by research leads [166, 257]. As such, research has called for participatory research to be careful in considering the proper applications of use [234], increasing the length of time and rigour with which participants are involved in the design process [270], and ensuring an intended impact that showcases all parties' expertise and knowledge in the final product over simply being ticking boxes over collaboration [195].

### 2.3.2 Focus Groups

In *A Practical Guide to Focus Group Research*, Breen prompts the reader to ask "*what do I ultimately expect to get out of this research?*" - when the answer involves perspectives, insight into experiences or testing ideas on domain experts, she advocates for a focus group as the key method [56]. Focus grouping is often used to understand user perspectives through group communication and collaboration [209], with particular utility for examining users who work in group settings, as it can help to extract collaborative activities through iterative and sustained discussion [190]. The method has been shown to be effective for highlighting to researchers particular areas within a domain to focus upon [146], and participatory analysis and self reflection activities within group research studies have been demonstrated to improve engagement between domain experts and researchers [77]. As with participatory research in general, it has been noted that there can be problems with power dynamics when participating in co-design activities in focus groups, and as such participants should be given control and authority when given activities, something we note to do in our own work [11].

### 2.3.3 Reflective Study

Utilising personal and professional reflection as a method of study emerged out the 3rd wave of HCI, and refers to theoretical critical engagement with research with an aim to improve the standards of study carried out in the future [29, 262, 263]. Dourish claims that, as researchers, engaging in reflection over good design, appropriate methodologies of research, and analysing data can make HCI a stronger and more stable genre [103]. Reflective design has become popular in recent years with practitioners placing a critical lens on ethical planning of studies [72], design presentation to user groups [97], sustainability in device construction and study [211] and prototype deployment [296]. Reflective practice as a primary method in research has been the subject of debate over its efficacy and necessity [99], but often helps when engaging with marginalised or sensitive groups. [148, 153, 181].

### 2.3.4 Design Fiction

Design Fiction is a method of speculative research from the early 2000s [182] with a focus around combining specialist experience and knowledge with imagination and creativity to produce ideas of where the future of technology may lie. HCI researchers have utilised DF for a variety of scenarios, such as where directions and trends of research will be heading [27], or highlight potential ethical issues of emerging technologies [28]. Whilst it has attracted controversy as a sole method of study, it has been seen as a useful method in combination with other methodologies in contexts such as working in interdisciplinary spaces where a shared research language has not been established [210], bringing together researcher and participant perspectives in a manageable way [76] and teaching non-experts about HCI and CSCW fundamental principles [246]. In short, DF is utilised best as part of a toolkit in certain circumstances to aide in reflection and extracting tacit knowledge [30].

As such, DF holds value in medical contexts of interdisciplinary research - Tsekleves et al utilised DF alongside co-design activites to imagine improved devices for deployment with older patients [301], Stead et al utilised DF in their proposal for DIY-based methods of designing personalised medical devices [282] and Strachan and York have both produced case studies that demonstrate the efficacy of DF in healthcare for producing plausible and realistic ideas that can be adopted in the near future [288, 322]. Here, we see that current research has highlighted the validity and utility of adopting DF as one arm of a multi-faceted study when dealing with domain experts in a context like healthcare as a

method of bringing together experience, knowledge and creativity to produce an insight into where participants believe design efforts may (and should) be focused in the future.

### 2.3.5 Working With Clinicians

Utilising active practitioners as participants in independent and "outsider" research studies (i.e. research that the practitioner has no ownership or relationship to) is recognised to present challenges, but there is significant variation in the reasons why and the best ways to address and rectify it [239]. The most common (and, to an extent, understandable) reasons provided include a lack of available time to spend participating, poor financial incentives, and a physical and cultural separation from the traditional academic community [113, 242]. A survey of 1050 American gastroenterologists concluded that a lack of financial and professional incentives was the primary barrier to voluntary participation in academic, factors addressed by industrial research, preferred for its lack of complex protocols, shorter duration and better compensation [100].

There are indications, however, that there are deeper factors than simply a lack of monetary compensation or a lack of free time; Baathe claims that physicians tend not to engage in early stage research because it offers a lack of personal and professional fulfilment - practitioners desired to feel educated and that their work will improve following participation [63]. Kreindler's work on collaborative healthcare work supports this, finding that private physicians declined to partner in research projects that they felt were *inauthentic*, when offers were transactional rather than cooperative [171], something often found in HCI research; Singh claims HCI tends to bring its research *to* clinicians, rather than offering collaboration [271]. These findings imply that physicians can be motivated to participate in outside research if it offers tangible benefits to their careers, even if more materialistic incentives are not offered. Alternative methods including gamification of participation and addressing research requests from participants in the planning process have also been effective [286, 142], demonstrating that increasing active physician engagement in research can be achieved through promoting professional fulfilment, and approaching participation in research as a collaborative and interactive process.

### 2.3.6 Research Gap

What we have demonstrated in this section is the utility of methods employed when investigating perspective-based research, and the methods of involving clinicians as researchers despite the difficulty in recruiting active practitioners for participation. What appears vacant, however, is employing these methods alongside one another; focus groups, reflexivity and design fiction have been employed in medical contexts [288, 51] but often without leveraging the existing knowledge over best practices to recruit and utilise clinical perspectives, and certainly not in a radiological context.

As such, there is room for work to be conducted with radiologists as participant-researchers, utilising their tacit knowledge and experience in *reflective* and *speculative* study whilst also being able to conduct traditional assessments of practice. There are clearly caveats to using practitioners in collaborative situations with other members of their community as demonstrated by Donetto [101], but utilising practitioners as collaborators with the *research team* such as in a reflective environment remains untested.

## 2.4 Diagnostic Medicine

*"The work of the modern radiologist is multi-faceted"*
- Brady et al *The Role Of The Radiologist In The Changing Word Of Healthcare* (p. 5) 2022 [52]

Here, we provide a brief overview of the current state of digitised radiology and the debates within the community, before examining some areas that are specific to each chapter about the psychology and design of tools and interventions before probing the lack of holistic understanding of work in-situ.

### 2.4.1 Technological Interventions in Radiology

Radiology has always been, almost by definition, a technologically reliant discipline. Thomas and Banerjee stake the beginnings of contemporary radiology at 1971, with the introduction of the commercially available Computed Tomography scanner, followed by the Magnetic Resonance Imaging Scanner in 1978, and finalising with Hammersmith Hospital becoming fully filmless in 1996, when the Picture Archive Communications System was introduced, effectively digitising the radiological workflow of the whole hospital [298]. PAC systems were continuously controversial in the 2000s, with a recognised increase in

autonomy and efficiency but an accompanying significant rise in workloads [175, 144], but have since become the industry standard.

The modern debate regarding technology in radiology surrounds the advent of Artificial Intelligence systems. With applications in image segmentation and analysis, 3D reconstruction, decision support tools, administrative automation and more, here has been a stream of medically authored publications advocating for AI integration into workflow [139, 169, 206, 289, 310, 320, 126, 59, 158]. ML and AI programs have been shown in laboratory experiments to improve speed of x-ray processing, identifying intracranial haemorrhages, centralising and cleaning health data, and "clearing the noise" of scans to enable radiologists to more easily identify lesions [94, 35, 15, 22].

Due to the fact that the primary output of a radiologist is their written diagnostic opinion, natural language processing applications have also been a common endeavour to extract key clinical information and generate actionable outcomes [92, 302, 117, 116, 276, 119, 274, 143, 290, 159]. However, a common critique of these efforts have been that they are primarily 1) utilising corpus's of pre-written reports to run tests on and 2) in small sample, laboratory applications. There has been little research on applying these systems in the wild, and literature has forecasted that they are likely to be inefficient due to their need for a standardised radiological lexicon, something that juxtaposes with the free-text and personalised nature of current reporting standards [237].

The introduction of these machine learning and AI based systems has caused significant debate within the medical diagnostics community. Chockley and Emmanuel's *The End Of Radiology* [78] identifies 3 key "threats" to the discipline - de-institutionalisation, a lack of payment reform, and the "ultimate threat" of machine learning. They cite factors including the lack of accountability regarding diagnostic errors, algorithmic accuracy compared to human performance and the increasing ubiquity of these systems in medical environments as the reasons for practitioners to be weary of the rise in medically focused machine learning [78]. The concern over responsibility in misdiagnosis is reflected in work by Neri [219], and other factors expanded upon by Wong et al, who believe that whilst job replacement in clinical situations is unlikely, the fear and anxiety over it comes from genuine concern and should not be dismissed out of hand [319].

The sensationalist nature of the title is emblematic of the current environment of medical AI, but Chockley and Emmanuel's views are controversial: Chan and Siegel's *Will Machine Learning End The Viability of Radiology as a Thriving Medical Speciality* provides a riposte to *The End of Radiology* by opining that automating tasks will be useful to aiding the human radiologist, but that the human-in-the-loop will always be required for trustworthy analysis and sign offs [73]. This perspective is echoed by similarly tabloid titles such as Recht and Bryan's *Artificial Intelligence: Threat or Boon to Radiologists* [249] and Pakdemirli's *AI in Radiology: Friend or Foe* [230]. Despite titles one may consider "clickbait", the prevailing opinion in these publications is that AI should be considered a *tool* to assist the practitioner in the same way a mouse or a scalpel is a tool to aide, but it is evident that there is a need to quell the anxiety among radiologists of being replaced. This goes alongside the growing concern regarding the lack of proper digital literacy among practitioners: studies in Vietnam, Ghana and Pakistan found that, despite ambivalence or even positivity towards novel AI implementations, levels of knowledge and education around digital systems and technology was low [300, 108, 5].

### 2.4.2 Communication In Health

Communication is, naturally, a vital aspect of triaging, diagnosing and treating a patient [273]. In radiology, where the practitioner's primary product is their medical opinion, communication is the *raison d'etre* of the radiologist [174], with emergency radiologists spending upwards of 10% of their whole shift simply communicating results to referring clinicians [229]. Well constructed and accurate communication holds extreme significance to the discipline: the failure to rapidly communicate an actionable finding has tangible detriment to the treatment of a patient [20], while 38% of communication errors in radiology result in a sentinel event [268], and a significant amount of lawsuits against radiologists in the USA cite a communication breakdown as a contributing factor [16].

Designing for and implementing methods of more effective communication, however, can prove difficult. Communicating rich detail in a highly specific domain such as healthcare requires a high level of tacit and implicit knowledge, something that can be hard to accommodate for in UX design [279, 280]. The continuously evolving landscape of digital radiology make it difficult to implement longstanding digital interventions

[212], and the variety of ways in which communication can break down alongside simple communication of results (including ordering scans, scheduling patient interventions, and interpreting results) mean that a social approach is required alongside a technical one [212, 268]. Additional complexity arises when implemented methods clash between disciplines: standardised reports have been shown to be preferable to referring clinicians, improving the legibility and identification of actionable findings, but are widely disliked by radiologists themselves [157]. Sonographers also spend a considerable amount of time communicating with patients and conveying diagnostic results, but are trained in "good communication techniques" as part of their medical education, something not found in radiology [54]. Even within radiology, focus groups with Swedish radiologists found practitioners dislike the way their peers discuss results and medical opinions [115].

Some interventions have been proven to be effective - providing patients with copies of their diagnostic reports improves their levels of satisfaction with the care they received [203], and triaging requests and reports with digital or human translators reduced the number of rejected scans or reports in hospitals with high numbers of international practitioners [79, 109, 75]. Most importantly, the significance of "*being there*" and employing observation and cultural integration as methods of research have been demonstrated to be vital in understanding communication methods in demographics that employ high levels of tacit knowledge and domain expertise [10].

### 2.4.3 Interruption and Task Resumption

Psychological and sociological sciences have cohesively agreed that interruptions when conducting complex cognitive tasks have a detrimental effect on the end result. These can include an increased lag in the time required to resume the task compared to one that is not cognitively complex [8], increased error rates and overall time complete tasks [176], and a decrease in the ability to recover an in-progress complex task leading to a higher likelihood it will be abandoned and re-started [9]. As such, there has also been a wealth of literature on how to address interruptions when conducting tasks in safety critical environments where error rates can have a serious effect. Generic work on task resumption has demonstrated that factors such as priming the user for an interruption, preparing them for disengagement, providing visual cues for resumption and timing an interruption for a lower cognitive part of a task can all lower error rates and the time required to complete

cognitive tasks [44, 185, 90, 13]. Domain specific task resumption has often focused on users in control of a dynamic environment that requires multitasking and split focus, such as in vehicles or planes where distractions cannot be easily dismissed [177, 218, 205, 48, 129, 111].

Due to the nature of healthcare environments, interruptions are accepted as necessary and unavoidable [254], but interruption mitigation has been frequently studied - reducing interruptions to medical administrators improves the amount and quality of task completion in hospitals [67], and utilising cues for task resumption in emergency medical departments can reduce error rates [114]. Radiology has also been the specific topic of study for interruption mitigation, due to recognised the negative temporal and qualitative effects they can have on an overworked radiologist [266, 315, 216, 321]. Changing the location of a radiologist's office to make it less easily accessible to casual drop-ins, reconfiguring phone lines, remotely vetting potential interruptions and introducing members of staff to help ease the burden on requests have all been shown to be effective when implemented in diagnostic environments [311, 198, 23, 314, 145, 33, 34].

There are, however, problems with the existing research on interruptions in radiology - many of the proposed solutions require significant financial or infrastructural investment, something many hospitals cannot commit to. Additionally, there is evidence that the practical ill effects of interruptions in laboratory and study controlled environments do not always accurately translate to real world impacts, and there have been concerns that overly scientific and descriptive analysis of cognitive task resumption can miss the casual and sociological effects that interrupting highly complex tasks have on the medical ecosystem [152, 135]. As such, there are calls to conduct more holistic and in-the-wild assessments of work organisation and the role that disruption has on radiologist's work in-situ [254].

### 2.4.4 Voice Based Interactions

Voice-based Interactions in safety critical environments are common, as they allow for handsfree and naturalistic interactions to control devices and software in situations where concentration is needed but tools have to be manipulated [138]. In this manner, VUIs have been deployed in increasingly uncommon and specialised domains such as space [305] or industrial welding [19]. Medical environments have often been a venue for the application of VUIs, with, with almost half of all modern start ups involving a medical device utilising

some form of speech recognition [292]. Speech to text systems appeared in radiological contexts as early as the 1997, when Lai et al's *MedSpeak*, a system allowing for continuous speech recognition when dictating diagnostic opinions, was published in CHI [172].

Speech to text has become the standard in many Western radiology departments since then, and is recognised as the industrial standard in the NHS as the primary way to author reports on medical studies [265]. Literature on the efficacy of speech based dictation in radiology departments has been widely discussed, with most studies agreeing that 1) dictating to a VUI is faster than utilising a transcriptionist and 2) it produces more errors. Studies have variations in the categorisation and variation of errors, methodologies used to assess accuracy, and additional variables, but the generalisable findings remain consistent from the late 20th century to contemporary publications [6, 14, 26, 37, 74, 140, 136, 208, 244]. Considering the widespread adoption of speech recognition in modern radiological contexts, the implication is that the speed at which a practitioner is now able to produce a report is worth more than the potential mistakes that may come from it. This is reflected in a recent survey of radiologist's satisfaction with their speech recognition systems - 70% of clinicians were satisfied with their system, but the key factor in this was the efficiency it allowed for, not the accuracy of their work [130].

The controversy and divergence of opinion, however, comes in the evaluation of *why* and *how* mistakes are made, and their impact on the end-quality of treatment that a patient receives. To address the former, studies have demonstrated that users need high levels of support and assistive cues when using VUIs, and that simply providing a microphone and speaker and assuming intuitiveness will not suffice - natural adjustments such as hyperarticulation when presented with misrecognition are common, but are not shown to be more effective at reducing error rates and can infact produce further errors when a user becomes frustrated with a VUI that does not accurately recognise their words [85, 283, 306, 214, 213, 215, 160, 127]. Additionally, when implemented in an existing workflow, effort must be made to ensure that a VUI suits contextual and pre-existing preferences, or else they are likely to fail [294].

In terms of the latter, there are growing concerns that VUIs in clinical situations should be re-assessed, due to the prevalence of potential errors each record can produce [1].

Further issues such as a perceived inability to accommodate for "foreign" accents [14] and implications that personal productivity actually *decreases* but can be hidden by wider increases in departmental gains [140] mean that, for some, *"the process of digital dictation is not applicable to the intended context of use"* in modern clinical settings [307]. There is a need to develop better models of interaction and a wider holistic mapping of the role that speech to text systems play in the authoring of radiology reports.

### 2.4.5 Research Gap

We have demonstrated here that most facets of radiological technology have, individually, been assessed in some way; the nature of task analysis, VUI systems, communicating important results and the growth of the digitised radiology department have all had multiple angles of research. This research is often not done in-the-wild due to the difficulty in gaining access to health environments and patient data, and as such often examines the "*perfect world*" of utilising these devices, systems and frameworks. It is, for example, understood that VUIs are quicker and more error-prone when conducting analysis on corpus reports, [26, 208, 37] but these studies offer us little in understanding *how and why* this may have its origins in technique or device design.

Our examination of WAD in HCI has revealed the importance of *being there* when tasks are conducted in a real-world environment, but most current research on radiological practice takes an isolationist view of device design and interaction rather than a holistic one. Work on communication has highlighted the need for frameworks to reflect existing concerns of clinicians, but this integration has not yet taken place. To understand the necessity, efficacy and efficiency of new technological advancements, work must be conducted using the methods explained and examined in the previous 2 sections - an area seemingly lacking in current investigations of radiology and devices.

## 2.5 Addressing Research Gaps

As proposed in the introduction to this chapter, we present the reader with a gap in research: radiologists are heavily technologically reliant, but can be overlooked in traditional HCI *Work As Done* interactional study due to their isolated "*hidden*" nature and lack of patient interaction. As such, little work exists that offers to reconstruct and

analyse the duties of the radiologist despite their central importance to the analysis and treatment of a patient's health. What is required to address the research gaps presented in this review is a series of studies that

1. Successfully integrates a Work As Done methodology into a workplace study of holistic practices in radiology departments
2. Utilises existing knowledge of participant-led study alongside strategies for involving clinical participants to successfully conduct assessments of current and speculative devices for implementation in radiology departments
3. Provides *actionable* and *tangible* outputs from this Work As Done lens that practitioners, designers and managers can utilise in their everyday tasks

Accordingly, this thesis sits across the Venn diagram of Work As Done, Participatory and Collaborative Research and Diagnostic Medicine to form a unique examination of radiology from an HCI standpoint. The following chapter will examine the methods and approaches adopted throughout the thesis that aims to address these research gaps provided and present work that meaningfully contributes to existing knowledge and understanding of the radiological and health HCI domains.

# Chapter 3

# Approach

## 3.1 Outline

Here, we present our main approach and methods when conducting the research that constitutes this thesis. We explore our methodological choices and their background, the scope of our focus, and the practicalities that shaped the conduct of each study. The aim of this chapter is to provide an all-encompassing overview of the style and genre adopted throughout, allowing for a better reading experience but also to better situate the reader in the context in which we have undertaken our research.

We begin by establishing the focal points of the entire thesis: the radiologist. We then deconstruct our key focus and establish the semantics and purposes behind our research question, before exploring the methods utilised throughout to investigate it. Finally, we establish some constants that will appear throughout each study such as the settings, health boards, and other factors that are pragmatic and practical in nature to avoid continuous repetition and re-establishment.

## 3.2 Scope

### 3.2.1 Why Radiologists?

This thesis will chiefly examine the radiologist as its subject. We have already established in our Introduction chapter the role and duties of the radiologist and other diagnostic practitioners (including reporting radiographers), but it is important to establish why we

have decided to frame our studies around this particular discipline.

When considering digital device use and implementation, diagnostic radiologists are a unique subject to examine. They are not patient-facing, and primarily operate independently and in private. They perform their medical analysis and provide a clinical opinion and recommendations of treatment without having seen, met or conversed with the patient they are examining. They work across the hospital, taking medical studies from multiple departments and acting as a singular point of contact for picture analysis. These factors set them apart from clinicians who work on the wards, surgeons, general practitioners, anaesthesiologists and other medical doctors. The ways in which a radiologist work in a hospital provides an insight into the behind the scenes of the patient's journey, the aspects that are not visible [250].

The fact that a radiologist is a different kind of subject to examine is interesting in itself, but beyond this our reason for focusing on this type of clinician is their relationship to the digitisation process and novel implementations of technology. The reporting radiologist's workflow is now almost entirely digitised – the scans produced are digital, uploaded to a cloud-based server and downloaded onto a personal computer, the report is produced through automatic speech-to-text programs, the image is analysed and manipulated with the aid of machine learning algorithms, and the final product is exported back to a digital cloud-based server.

This means that radiologists are also often at the forefront of the digitisation process. The first Picture Archive Communication Systems (PACS) and Radiology Information Systems (RIS) were implemented at the turn of the century, meaning radiology departments were among the first in the hospital to be almost entirely digital [173]. Now, with the advent of Artificial Intelligence (AI) for decision-making in clinical contexts, radiology is often the prime area for initial implementation. It's place as the digital focal point of the hospital is well deserved, but there has been little work on the base level of how these practitioners interact with digital systems when working.

The fact that the radiologist is entirely reliant on digital systems, and does not interact with the patient themselves, makes them an ideal candidate for interactional study.

Interaction with digital tools and devices constitutes the vast amount of their work - there have, however, been few dedicated studies into the ways in which radiologists utilise these tools when constructing and authoring reports [7]. Similarly, the growing calls for AI implementation in radiology departments necessitates an understanding of existing workflow and practice. As such, we frame our investigations on radiologists to address this.

### 3.2.2 Research Focus

The key focus of this thesis is on device use when constructing radiology reports holistically, in-the-wild and with an emphasis on ethnomethodologically informed ethnographic study. By framing these key terms and explicitly outlining our focus, we hope to align the reader to what we will (and will not) be covering in the course of the following chapters. Here, we will briefly construct this sentence to explain and define these terms.

We use the phrase device use when constructing radiology reports to refer to the interactional process of a practitioner receiving, analysing, and authoring a clinical opinion on a medical study. This process brings the practitioner into contact with a variety of hardware devices and computational peripherals, as well as dedicated medical and 3rd party software. This includes purpose-built items such as dictaphones and image scanning programs, as well as generic tools implemented in a specific domain like internal telephones and instant messaging apps. It is worth noting that the British Government define software and apps as a "device" for regulation and procurement purposes [131]. We, in turn, define all of these devices as safety critical – whilst they are not directly involved in administering treatment to a patient (as say, a radiotherapy or MRI machine would), their misuse or significant failure would have a significant impact on the patient journey. Software failures could result in the wrong patient's data being reported upon, or mis-transcription that produces an incorrect report; hardware failures could prevent a reporter from being able to communicate an emergency actionable finding. In this setting, everything available to the practitioner contributes to the safety critical setting [16].

Under this scope, we are only examining the reporting process – we do not venture into the x-ray room or the surgical ward to examine radiographers and radiologists conducting interventional procedures, carrying out scans, or communicating with patients at their bedside. This means that our scope of device use does not extend to MRI or CT

scanners, for example, instead allowing us to focus on the individual radiologist in their personal setting. There were practical reasons for this choice, as extending this scope would bring us into considerably more interactions with patients, a difficult ethical area to navigate as part of doctoral studies in the NHS, as well as methodological choices – the reporting process has not been the subject of scrutiny in the same way the scanning and treatment process has.

In line with this, our scope concerns holistic and in-the-wild study of these devices. Chapters 4 and 7 examine the diagnostic workflow and timelines, and Chapters 5 and 6 explore interactive processes with an eye on reconstructing as a whole; we do not take a singular device and subject it to usability testing, instead focusing on the ways in which these devices come together to constitute an 'interaction'. Whilst we examine dictaphones, image manipulation software and the other devices used when constructing a report, these is an emphasis on how they are used in their finished state, as part of how they are implemented in a real-world space. We are not applying formal methods and mathematical analysis to each device to determine usability, instead applying a wide lens to view the process as the user would.

Finally, we adopt an ethnomethodologically informed ethnographic approach – we will discuss the techniques and methods specifically chosen below, but for the scope of our research this means we have chosen to conduct primarily people-based study. We have created facsimile devices (devices that look and act as the real device would, but adapted for lab use) to test and presented designs to participants, but there are no interactional devices deployed. This allows us to focus on the interactional process as they are in the real world avoiding designing and implementing probes that would be inappropriate for such a highly developed and specific domain. The key data being created and examined for this thesis comes in the form of observable and repeated actions that constitute a behaviour, subjected to analysis to form a scientific understanding of the interactive processes and preferences of this user group.

## 3.3 Methods

### 3.3.1 Ethnography, Ethnomethodology and the Research Lens

Here, we will briefly explore the background and context of ethnographic and ethnomethodological approaches, before examining the lens under which we focus our studies.

Ethnography is an overarching term referring to the study of culture, people and life. It is anthropomorphic in history and nature, with the generally agreed upon origins in the work of Malinowski – his 1922 publication Argonauts of The Western Pacific, examining the lives of natives in New Guinea, is often recognised as the seminal piece of ethnographic research [193]. Whilst the phrase is now considered to be apocryphally attributed, Malinowski's method encourages the researcher to "come down off the veranda", to leave the comforts and settings of the academic office and observe participants in the wild, and immerse oneself in the culture they are studying. As such, ethnography places a significant emphasis on the practice of observation, documenting the real and accountable activities of the subject in question.

Ethnography has become a particular method of note in Human Computer Interaction for its ability to shape design – the "turn to the social" movement of the 1980s and 1990s highlighted the potential of ethnographic study to supplement system design and implementation beyond what designers believed would be the most efficient method. Suchman's work in the mid 1980s pioneered the "workplace study", a method of observing users in their natural environment interacting with technology (in this case, a Xerox photocopier) [291]. Suchman's use of ethnographically inspired methods became a template for other HCI examinations of work in-situ, analysing technological use and interaction "in the wild". In the words of Crabtree, "The main virtue of ethnography lies in its ability to make visible the real-world sociality of a setting" [86].

Ethnomethodology was initially proposed by Garfinkel in the 1967 publication Studies in Ethnomethodology, and concerns itself with the "common sense" of human behaviour and social order [220]. In Button's summarisation, "*Ethnomethodology is not a 'theory', nor is it a 'method', instead it is a way of doing social science*" [61]. The practice is indifferent in nature, meaning it aims to build social structures through an observable, data-driven

"bottom-up" approach not based upon prior understanding or beliefs about the topic in question. Ethnomethodology focuses on the mundane, the invisible and the background of social actions, situations and structures to address the "*invisibility of common sense*" [245].

Ethnomethodologically Informed Ethnography (EIE) came out of the rejection for traditional ethnography as a method of system design – Crabtree's Ethnography Considered Harmful is a prime example of HCI practitioners seeking to highlight the ways in which ethnography as a practice had regressed to overly simplistic and generalist conclusions with little relation to practice system design [87]. EIE is the "*study of people in practical action*" [245] - it focuses on Work As Done over Work As Imagined (later unpacked in Chapter 3). By basing our ethnographic investigations on an ethnomethodological framework, we allow for a more focused and constructive set of conclusions to be made from our data. It prevents the work from becoming overly "fluffy", and provides a detailed look at the practical actions of our participants when interacting with pre-designed systems.

To summarise, our work takes the underlying notions provided by Ethnomethodology, a focus on the unseen and invisible social interactions that constitute "common sense", and uses it to frame our ethnographic inquiries into practice. This provides us with a structured and analytic look at the radiologist at work.

### 3.3.2 Data Gathering

Throughout this thesis, we utilise a series of methods of data gathering including non-interactive observations, focus groups, laboratory studies and interviews. These methods are primarily human-centred and ethnographic in nature, and serve to reconstruct behaviour and interaction in a way that allows for proper analysis and discussion. The goal of these studies is to examine the actions, events, goals and settings of practitioners beyond simply recreating "scenic descriptions" (simply describing what we see with no critical engagement [124]), instead providing the reader with a scientific understanding of what constitutes a practitioner's interaction with the tools at their disposal. This requires an understanding of the context in which the practitioner has learned their skills, their tacit medical knowledge, their opinions and perspectives on the efficacy of these systems, and a wider foundational understanding of the ways in which the hospital social structure operates. As such, throughout this thesis we adopt a variety of methods to examine all facets

of the radiologist at work, discussing with them directly in groups or individually and observing them independently in different contexts. Many of these methods examine the same interactional process from contrasting viewpoints, constructing a multi-dimensional representation that incorporates the user's perspectives with our own as researchers.

Chapters 4 and 7 utilise a heavily participant-led method of allowing practitioners to voice their perspectives, and using this to shape our conclusions. These chapters take the form of semi-structures focus groups, a way of gaining tacit knowledge from expert users to contextualise our other observational findings. Chapters 5 and 6 are more traditional ethnographies, and form the basis of most of our contributions – they take place in both natural and constructed environments, but the primary method is non-interventional observation. We allow (and encourage, as much as is possible) participants to act as they would if we were not studying them, and our data comes in the form of descriptions of their actions.

### 3.3.3 Analysis

Ethnographic data is, by its nature, "Eclectic" [87]. As the majority of our data is qualitative, we primarily utilised thematic methods of analysis that focused around transcription-based examination. In Chapters 4, 7 and 8, the majority of data gathered is the spoken word, and this also forms a significant portion of the supplementary data in Chapters 5 and 6 – even in these, our data for analysis consists of note or discussion based on observable behaviour. As such, we choose to adopt ways of analysing and presenting that are more suited to a participant's voice and perspective than rigid quantitative data. Throughout, our goal is not to demonstrate data in a numeric way, although we will occasionally refer to certain statistics and figures to illustrate a key finding or to add "colour" to a thematic element. Instead, we will often present data in the form of thick descriptions, vignettes or fragments of activity, and illustrative commentary with participant's direct quotes; this allows us to explore a more personal and lived experience of the user group we're studying.

### 3.3.4 Participants

In Chapters 4, 5, and 6 our participant pool is made entirely of clinical practitioners. Chapter 7 also incorporates technical managers, but the vast majority of our work centres around examining participants who are employed by the NHS. It is worth establishing who these participants are, and their roles within the patient treatment journey. We have

already discussed "the radiologist", as they are the primary focus of this thesis, but it is important to note the differences between this wider definition and the individual groups beneath its umbrella terminology. The NHS is an integrated and nationalised health service throughout the United Kingdom, and as such many of our practitioners work across multiple sites and hospitals through the year. NHS Wales has 6 University Health Boards (UHBs) and 1 Teaching Health Board of Powys, and our research is based primarily at Cwm Taf Morgannwg UHB as this is the board that governs the NIAW. However, we also have participants from Swansea Bay and Cardiff and Vale boards in Chapters 4, 5 and 7.

Our participant pool includes Consultant Radiologists, Trainees, and Reporting Radiographers [221]. All of these practitioners analyse medical studies to produce a diagnostic opinion, but vary in their experience and the format of their report. Consultants are fully qualified practitioners, having completed their training and specialisation – they analyse images independently, and with their medical training often provide treatment advice or recommendations to the referring clinician, resulting in a report that has a mixture of evident findings and expert medical opinion.

Trainees are qualified doctors who are undergoing their specialisation; the training process takes between 5 to 7 years, including a foundational phase and speciality training. These trainees will often report on real patient images, but will require their work to be checked by a consultant before it is confirmed to be accurate and appropriate. Much like consultants, these trainees will utilise their medical training to provide treatment advice, but this is less the focal point than providing sound and accurate identification of actionable findings in a scan.

Radiographers are the technicians and practitioners who operate the scanning technology required to produce medical studies. There are a variety of roles under the "radiographer" job description, but primarily in this thesis we work with Diagnostic Radiographers. These are technical specialists who have undergone several years of dedicated training in radiography and report on medical studies, but do not have a medical degree. As such, they learn how to analyse for actionable findings, but often do not provide treatment or clinical advice. Instead, their reports read considerably more matter of factly, demonstrating to the referring clinician what can be seen on the scan and allowing

them to make the decision of treatment. Junior radiographers just beginning in their training will have all of their reports proofread for accuracy, but more senior reporting radiographers operate much like consultants, working independently and liaising directly with the referring clinician.

These participants are often treated in our work interchangeably – they all utilise the same equipment, on the same software, and work out of the same spaces. However, the content of their reports can vary, as can the ways in which they interact in the social structure of the hospital. It is important to note that when we refer to "radiologists", we are using this as a blanket term to refer to diagnostic practitioners who construct reports, unless specified.

### 3.3.5 Conducting Fieldwork and The Position of The Author

In all of the studies found throughout this thesis, the work was conducted face to face, almost entirely in-person and solely by the author. This means that in every aspect of data gathering (aside from in Chapters 6 and 8, where some work was conducted face-to-face but via teleconferencing call) the author was in the room with the participant when they were taking part in the study. As such, we must address the impact that the author being in attendance has on participants - this is often referred to as the *Hawthorne Effect* [192], but here we will refer to it as observation bias. The effect refers to the supposed phenomenon that participants will change their behaviour in research when being observed, and is often used as a caveat when discussing findings of pure observation research [192]. In modern ethnographic work, observation bias has become an accepted part of the nature of recording naturalistic and human-centred data, and some have even gone so far as to dismiss the negative impacts of observation bias as myth, fallacy or a useful dimension of ethnographic study [204]. In HCI, discussing observation bias can be controversial due to the prevalence and necessity of recording data of interactions either through human observation or through camera recording [224] - previous work has established that whilst having a member of the research team observe a participant can have an effect, it can be less prevalent than task experience or the medium of the task itself [295]. Additionally, whilst participant bias towards the researcher in HCI was been documented before when presenting prototypes to participants [97], in all of these situations bar Chapter 7 we were not presenting designs to participants and instead asking them to use the devices and tools that they were already familiar with, and so we will examine this directly in Chapter

7. Finally, and contextually specifically, radiologists in the NHS train through observing senior members of their team and by being observed when they perform their tasks as trainees [241], meaning they are familiar and experienced in being observed as part of their normal day-to-day routine. As such, we will not significantly caveat our work by leveraging observation bias over the validity of our findings, but we will acknowledge the existence of observer bias as an aspect of conducting ethnographic research.

Another dimension of conducting this work in-person and face to face, however, is acknowledging the relationship between the researcher and the participant and how this can affect the nature of the results we find. In every chapter, we are conducting work in collaboration with (and to an extent on behalf of) the National Imaging Academy Wales, meaning all work would be signed off by both NHS review and approval from the Director of the Academy as appropriate - this provided us with an element of authority, and meant that participants had both trust in the author in the room and thorough understanding (through information sheets and through discussion with the author and the Director of the Academy) of the nature of the work they were undertaking. As such, when conducting studies the author was treated as a peer when discussing elements of friction and failure, with participants eager to discuss their experiences and explore systemic and social dimensions of being a radiologist, an aspect that will be discussed extensively in the *Discussion* chapter. Further, due to the *opt-in* nature of participation in these studies, participants who may have felt that the research was not worth their time or did not want to participate simply did not take part, and as such we did not experience significant resistance from any of our participants that would have made data gathering difficult.

## 3.4 Settings

### 3.4.1 National Imaging Academy Wales

As the primary focus of our research is the radiologist's natural behaviour in-situ, we often utilised a clinical setting as our base of operations. In our collaboration with the NIAW, we frequently used their training campus as a space in which we could conduct studies, providing access to practitioners that would traditionally be difficult to recruit away from their work.

The National Imaging Academy Wales, located in Bridgend, opened in 2018 as part of an effort to increase the number of radiologists in the NHS. It delivers the Royal College of Radiologist's training curriculum whilst also admitting a small number of patients for treatment, making it both a training site and active medical practice. The campus holds regularly events in addition to its teaching, making it an ideal area to situate ourselves in when recruiting for and conducting studies.

Because of our ability to "set up shop" in the NIAW meeting rooms, we adopted a right place right time approach to recruiting participants [39]. For Chapters 4, 6 and 7, equipment would be set up in a room on NIAW campus on days when the Director of the Academy knew there would be an abundance of trainees or teachers for us to work with. The study information sheets would be circulated and advertised in advance, but with a continuously set-up room - the goal was to be flexible around an inflexible participant pool, and allowed us to maximise participation in our studies.

In some cases, such as that of Chapter 7, where large groups of participants were needed simultaneously, a date would be chosen that aligned with a reason for practitioners to be on NIAW campus, such as a guest lecture. We would then be able to ask participants to join our study, instead of inviting them to the University Campus or organising a dedicated date. Access to the NIAW site enabled this methodology and approach to recruiting participants, and as such became the main setting for the majority of these studies. It provided additional benefit by being a site familiar to participants, and so encouraged them to be comfortable when taking part.

### 3.4.2 NHS Hospitals

In Chapter 5, we adopted a different approach by situating ourselves in NHS hospitals. This was required to understand natural behaviour in-the-wild that we felt would be difficult to obtain on the NIAW campus. We used hospital sites from Cwm Taf Morgannwg and Cardiff and Vale for this study, and it marks the only situation where we operated outside of the NIAW building. Chapter 5 discusses the hospital sites in detail.

## 3.5 Ethics

All chapters of this thesis were approved by Swansea University's Research Governance, with Chapters 4 and 7 also receiving approval from the NHS under the service evaluation framework. For Chapters 5 and 6, we had to obtain full NHS ethical approval in the form of an Integrated Research Application System (IRAS) application. Chapter 5 holds IRAS approval number 313272 and Chapter 6 331882 [264]. IRAS forms were required for us to utilise clinicians and patient data as focal points for research, and mandated approval from the NHS Research Ethics Committee. These applications required documentation from Swansea University and from the relevant health boards to be approved in multiple formats simultaneously, something that presented a significant challenge – NHS REC approval has a reputation for being difficult to procure, and our experiences with this ethics system provided wor for our Discussion, where we examine the role that administrative difficulties can play in conducting research as an Early Career Researcher.

## 3.6 MSc Project and Groundwork

Prior to conducting the studies that constitute this thesis, the author completed an MSc project examining Trainee Radiologist's Perspectives on Machine Learning and Artificial Intelligence [80]. This initial piece of research provided us with a groundwork of understanding of the levels of digital literacy and knowledge over AI systems from both the trainees who participated and our stakeholder management team when organising the project. The results of this MSc project showed that the younger generation of practitioners were *enthusiastic* about the prospects of an AI-powered radiology department, but that levels of digital literacy surrounding what AI could do and what systems were already present were *low*. This gave us motivation to investigate what happens in radiology departments before attempting to implement designs and probes.

## 3.7 Conclusion

This chapter has established our main research scope and the methodology we adopt to address it. We provide a brief "whistle stop tour" of the topics and concepts that will feature heavily in the following chapters. By outlining these factors, we hope to situate the reader and orient them to the specific domain in which we are focusing. We outline

the specific area in which we are situating ourselves, with contextual descriptions of the participants and the settings in which they work, alongside a brief explanation of the ethnomethodologically informed ethnography method used in the following chapters as our research lens. Following this establishment, we will engage in an in-depth literature and background review that will further allow the reader to understand the research space in which we are operating, and establish the scientific niche to which we contribute our findings and discussion.

# Chapter 4

# Active Practitioner Perspectives On Digital Workflow

## 4.1 Introduction

Prior to undertaking studies in-situ, we conducted a study that allowed us to understand the environment in which clinicians operate to better situate the ethnographies that will follow in Chapter 5 and Chapter 6. By conducting a participant-led activity that focuses on the more bird's eye view of overall practice within the hospital, we gain a level of knowledge and context that will allow us to examine specific practices of radiologists in greater depth. This chapter consists of two iterations of a focus group conducted with a total nine active practitioners, working to establish insider perspectives on where friction and failure exists along the patient-clinician timeline of interactions. By asking participants to self-identify where they feel problems are, we provide ourselves with a level of competency and contextual understanding regarding the environment we will be working in, as well as establishing a comparison point for following studies.

Our focus groups establish a self-reported and group-codified timeline of patient interactions complete with areas of failure and frustration, which we then use to drive Design Fiction activities resulting in a "*Wish List*" of things practitioners believe would address their problems. This presents an representation of how the clinician perceives their interactions with digital systems, their levels of digital literacy, and user beliefs on system and device design – all of this culminates in a series of takeaways that we can use to compare

and contrast against when evaluating tool design and interaction in-situ. We present 3 key elements that participants would gravitate towards when discussing prevalent issues and how they wish to solve them of *Communication*, *Integration* and *Support and Infrastructure*. These elements serve as focal points to align ourselves when conducting our next steps.

In Chapters 2 and 3, we establish the need for a better understanding of *Work As Done*, meaning the real-world actions and interactions that constitute behaviour when operating in a healthcare space. To do this, we must understand the healthcare worker, the healthcare environment, and the contextual and tacit information that cannot be *seen* when tasks are being communicated and executed. Early explorations of digital implementation in medical contexts showed the high levels of failure that can occur when clinical understanding is not taken into account, with as high as 75% of medical IT projects failing before full integration in the early 2000s [184]. This lack of integration between healthcare professional and developer was described as still "*waiting for Godot*"; the mismatch between two highly specialised groups without fundamental understand of what the other was trying to achieve would always result in a lackluster system being implemented [313].

Utilising clinical-facing practitioners in ethnographic research as a method of orienting research focus to real-world areas of concern is a widely trodden area of literature, often used by institutions such as University Hospitals and nationalised health services as a way of internally investigating areas of friction [233, 46, 43, 89, 207]. Wong et al discuss the importance of including practitioner perspectives in the mapping of clinical workflow – they term this *mythical vs actual*, but the method aligns with our *work as done* approach [318]. Similarly, Bowen emphasises the necessity for stakeholder involvement when designing for implementation in hospitals – the complexity of the environment means that including practitioners in scoping exercises greatly increases the accuracy of future projects [51].

Here, we follow on in the tradition of conducting a participant-led exercise to investigate self-reported interactions with patients and devices but with a particular focus of intending the results of this focus group to guide us on an area in which to focus our research goals

for the rest of the project. Simply put, we are asking clinical participants to inform us of where they feel our research efforts should be spent.

## 4.2 Method

### 4.2.1 Approach

Due to the fact that the researcher had little knowledge of how the hospital was organised internally, it was deemed inappropriate to attempt to conduct research without an appropriate level of understanding and context from the radiologist's perspective. To gain this, and guide our further research, we decided to conduct a focus group, as we felt this would allow us to gain multiple perspectives at once that were, in themselves, subject to group consensus before they were provided to the researcher - interviews would require extensive cross examination and are sensitive to personalised experiences, whereas a study of this design would ensure that the common experiences are the ones provided as data, and these are the prevelant areas of friction that we are aiming to address with our research. Primarily, we based our approach to the planning, implementation and analysis of this focus group on Stewart and Shamdasani's *Focus Groups: Theory and Practice* [285, 284]and Breen's *Practical Design for Focus Groups* [56].

Steward and Shamdasani's textbook on conducting focus group-based research is an all-encompassing guide that was frequently referenced prior to and after our focus groups were conducted for inspiration on technique and stylistic choices. For example, as per the recommendations of the textbook, we adopted a conversational style with our participants that, whilst situated in their home ground, emphasised the researcher as an *outsider*curious to know about their experience, to ensure they were comfortable discussing what could be perceived as negative aspects of day-to-day workflow. Additionally, we ensured that our recording method was un-invasive, in the form of a small 4-track recorder that we pointed out to participants early, but was out of eyeline as so to not distract them. The influences of Stewart and Shamdasani's work can be found throughout this chapter and in Chapter 7, and served as the key piece of guidance for how we went about conducting ourselves when carrying out focus groups.

Additionally, by orienting ourselves in line with Breen's work we emphasised a discussion-based structure that hinged on the *collective experiences* of the participant pool. This meant that we did not enter the focus group with our own designs and factors to be evaluated by the participants, instead allowing them to construct their own materials and use that for a base of analysis and discussion throughout. As such, the content of the focus group was designed to be open-ended and participant driven as possible, allowing for participants to raise and discuss points within their group as opposed to with us as the research conductors. This left us with a lightweight framework and loose structure that could be pulled in any direction over the course of the focus group.

We also drew from previous collaborative and participant driven work conducted in NHS spaces. We refer to these in Chapters 1 and 2, but the *lessons learned* aspects that we established of conducting research with clinical participants instilled in us the necessity to make this an engaging task that diverged from what could be found in textbooks and seminars, to garner insight into what participants believed were the unmentioned and neglected aspects of practicing medicine under the NHS. This gave us the approach of having individualised activities that, again, hinged upon personal and collective experiences.

### 4.2.2 Participants

As outlined in Chapter 2, this work took place on NIAW campus. As such, we adopted the *right place, right time* methodology for recruiting participants [39]. We found a date on the training calendar that would result in a diverse pool of potential participants at the campus (in this case, a seminar series), and had a message circulate amongst attendees that we would be holding our focus group in a meeting room near where the seminars were being held.

We held two sessions, the first being with a group of five trainee radiologists from the same cohort in their 2$^{nd}$ year of radiology training, and the second being with a group of four consultant radiologists who knew each other but were not all from the same University Hospital Board. This allowed us for some aspects of "shared experiences" considering participants in all groups had British medical degrees and all worked in the radiological domain, but provided us with participants who were familiar with different tools and devices.

| Participant Number | Occupation | Focus Group Number |
|---|---|---|
| P1 | Trainee | 1 |
| P2 | Trainee | 1 |
| P3 | Trainee | 1 |
| P4 | Trainee | 1 |
| P5 | Trainee | 1 |
| P6 | Consultant | 2 |
| P7 | Consultant | 2 |
| P8 | Consultant | 2 |
| P9 | Consultant | 2 |

Table 4.1: Perspectives Focus Group Participant Information

Discussed anecdotally during the focus group, the trainee radiologists came from a variety of different sites across the UK and would use this to compare and contrast their previous experiences with patient interaction. Being trainees also meant that they were less experienced and familiar with the common systems and devices found in a radiology department, meaning they would be able to share details on entry-level usability and intuitiveness. The consultant group, on the other hand, had considerable amounts of experience and tacit knowledge regarding the interaction and treatment process, meaning they were able to speak on repeated and sustained issues that arose through years of practice. The combination of these two groups gave us a two-dimensional understanding of the patient treatment and radiology experience, as well as what participants would want to see as beginners in this space and as well-established practitioners.

### 4.2.3 Session Structure

In line with our framework from Breen, we split our session into 3 sections – an "opener" activity designed to get participants comfortable and discussing amongst themselves, a more heavily analytic activity to engage personal experiences and perspectives, and a final "lighter" more speculative activity that built upon the previous sections. These sections had a narrative connection, as we first asked participants to establish a timeline they were comfortable with, analyse and comment on it, then offer their own perspectives on how it could be improved upon in the future. As such, the content and data gathered of these focus groups was entirely shaped by participants.

*Part #1 – Establishing A Timeline*

After introducing ourselves and the nature of the study, and gaining consent from participants, we asked participants to construct a "timeline" of patient interaction from their personal and collective experiences. Here, we drew from Stewart et al's study on maximising clinician participation by *gamifying* various participatory activities – whilst Stewart's study used a competition framing, we utilised an approach that emphasised the shared experiences of participants by encouraging participants to create their own personal unique perspective on what the best timeline should look like, informing them that there would be a group comparison at the end [286].

We started by handing individual participants a stack of Post-It notes and pens, encouraging them to name as many phases of patient interaction as they could, not necessarily in order. Whilst we initially requested they operate independently, we attempted to give them as little guidance as possible in what would and wouldn't constitute patient interaction – as a result, participants would tend to *begin* independently before discussing with the group what they should or shouldn't include. Comments such as *"Oh I hadn't thought of that"* and *"I'm doing mine this way"* occurred frequently, demonstrating the collaborative end nature of the activity.

After constructing these timelines by themselves, we asked participants to place them in a linear order from first to last interaction – these timelines were then stuck to a large whiteboard for participants to examine collectively. We gave each group between 5 and 10 minutes to discuss which timeline was seen to be the most all-encompassing, allowing them to take stages from any and all of the timelines to construct the group's collective consensus. This left us with a finished horizontal list of Post-Its left on a whiteboard and offered participants a break, during which they could reflect upon the timeline they'd made.

*Part #2 – Identification of Failure and Friction*

Once consensus had been reached amongst the group of a comprehensive and all-encompassing timeline, we would provide participants with a different coloured set of Post-Its and ask them to identify common issues that arose during or between stages. This would comprise of each participant writing a factor on a note individually, standing up to

place it on the board, then discussing their reasoning behind it once they had returned to the table with other participants. This provided us with a full timeline that was annotated throughout with frequent pitfalls and areas of friction identified by participants. We would then engage participants in whether they thought the root cause of the function was as a result of, or could be improved by, a digital intervention. The aim of this discussion was to avoid particular emphasis on issues that participants felt could be improved with systemic or political intervention.

*Part #3 – Speculative Design Fiction: "Wish Lists"*

After examining and discussing the timeline and associated points of failure along it, we provided participants with a Design Fiction scenario: a fictional piece of software or digital device was being developed to address any of the problems that participants had identified in the previous activity. The participants would form part of the scoping committee for the design and functionality of this tool, so they had to provide us with a "*Wish List*" of things they desired this tool to do – we informed participants that there were no financial or physical limitations on the items in this list, but that they should be realistic for something that could potentially exist. This was informed by our MSc study, where our "no-limits" design fiction activity resulted in most participants inventing all-purpose robotic assistants [80].

Framing the *Wish List* in this way is built upon Slegers et al's exploration of utilising fictious scenarios in interventions for patients through design fiction [275]. Commonly, healthcare practitioners can struggle to feel involved as co-designers [51], but by utilising a scenario in which participants are involved in management and leadership we instilled a sense of ownership and stake in the project that gave them control over what functionality should be included, as opposed to simply providing ideas to a group leader.

This *Wish List* took the form of a series of items and bullet points that would be suggested by participants as part of a group, then once consensus amongst the whole cohort had been reached would be presented to the researcher. As such, it was an unstructured document that acted as a representative for what participants believed would "fix their

problems" as opposed to a finished design of a tool or system. After showing participants their finished lists and asking for final commentary or feedback, the session would end.

### 4.2.4 Conducting the Focus Group

The focus groups were conducted solely by the author in a room on NIAW campus. Prior to participating, participants would be informed of the author's relationship to the NIAW and the reason that the work was being held there - this also meant that participants did not treat the author solely as an "outsider"; it was known that the author was working collaboratively with the NIAW and the NHS in general, meaning the author was treated like a peer in the focus group space. Participants were sat around a round table in the middle of the room, with the author also taking a seat at the table but standing to move materials or write on the board. At the beginning of each session the author introduced themselves and the nature of the project, before informing participants of the recording device and placing it at the centre of the table. In this way, the author would also make it clear that they were operating from an academic and University-situated space, encouraging participants to be honest and reinforcing the anonymity of the study. By conversing with participants in this way, the author established themselves further as a peer and not an individual making an *assessment*, making discussion more conversational and allowing participants to reveal details about the political and social aspects of their work.

As previously mentioned, participants came from the same cohort (in the case of trainees) or health board (in the case of the consultants) meaning that they were familiar and friendly with each other both prior to and during the focus group session. This meant that the author had to do little in the way of encouraging participants, and all participants contributed equally to the session in terms of discussing among themselves.

## 4.3 Analysis

For the timelines and *Wish List*, participants' input was recorded directly into a digital spreadsheet to ensure items were not lost or re-assembled in the wrong order. In addition to a digital representation of both groups finished timelines (and data about extraneous steps that were not included in the final version) and *Wish Lists*, audio was recorded

through the entirety of each focus group.

This audio from the workshop was then transcribed by hand by the author before being thematically analysed by the author and an independent researcher (Ms Susannah Downie from Swansea University) who was not involved in the planning or execution of the study, and as such was seeing the data for the first time. Ms Downie also worked in a healthcare space, and as such had familiarity with the language used by participants that meant they were able to understand the content of the focus groups. They were provided with anonymised transcripts from the two workshops, and conducted coding independently alongside the author and discussion of their codes with the author to agree upon emergent themes from the workshops. Analysis of the transcripts was conducted in 6 phases inspired by Braun and Clarke's 2006 framework for thematic analysis [55] – both transcripts would be read in entirety by each researcher to familiarise themselves with the content. Next, an initial pass would be conducted where key phrases and quotes would be identified and tagged with commentary, then discussed between the two researchers. This discussion revolved around interesting findings and potential themes, before a 2$^{nd}$ pass of analysis was conducted with these potential themes in mind, codifying the initial codes into verified themes. These verified themes were then discussed a 2$^{nd}$ time by the researchers with a mind towards "seeing the big picture" and a resulting set of end themes presented here.

## 4.4 Results

In this section, we will present the participants' data without engagement or discussion, instead choosing to visually represent the results with commentary and contextual explanations. The reason for this is to tie our thematic analysis of transcripts in with our discussion element, as we found that quotations would be best served alongside broader discussion and thematic elements.

### 4.4.1 Timelines

Here, we will discuss the stages identified by participants, before moving to a codified timeline used to demonstrate areas of failure and friction. Across the two iterations of focus groups, participants identified a total of 30 unique stages of patient interaction. *Unique*

here refers to identifying a stage with particular terminology that did not overlap with another stage that may share the same content but was described or titled in a different way i.e. using different terms to identify the same process. The smallest list of stages comprised of 8 items, whilst the longest had 15 stages.

Initial and final stages were consistently agreed upon by participants, with *Discharge* occurring in 89% of all timelines, *Initial Testing* occurring in 78%, and *Clerking* in 67%. The timeline of interactions would often diverge into personal opinion after *Initial Testing*, with unique stages that only occurred once such as *Input from Health Centres* and *Allied Health Support* appearing alongside types of review, diagnosis or treatment. *Radiology* as a specific term only appeared twice, with most participants preferring to use generic language such as *Specialist*or *Additional* when referring to tests or treatment in order to include other disciplines. Of particular note is that, even prior to being asked to identify negative aspects of their interactions, some variation of the term *Waiting For Results* occurred in 56% of independent timelines, demonstrating how ubiquitous and unavoidable long delays were to the treatment process.

Figure 4.1 shows a timeline that was agreed upon by Group 1 to be "comprehensive", meaning it covered all aspects of the patient treatment process and they were satisfied with its layout and verbosity. It was constructed from several participants' individual timelines through discussion and collaboration, where they would approach the Post-Its and take from other timelines and re-arrange in real time until the group was in agreement. For this reason, this timeline will be the one utilised to discuss and illustrate our areas of failure and friction as a point of reference.

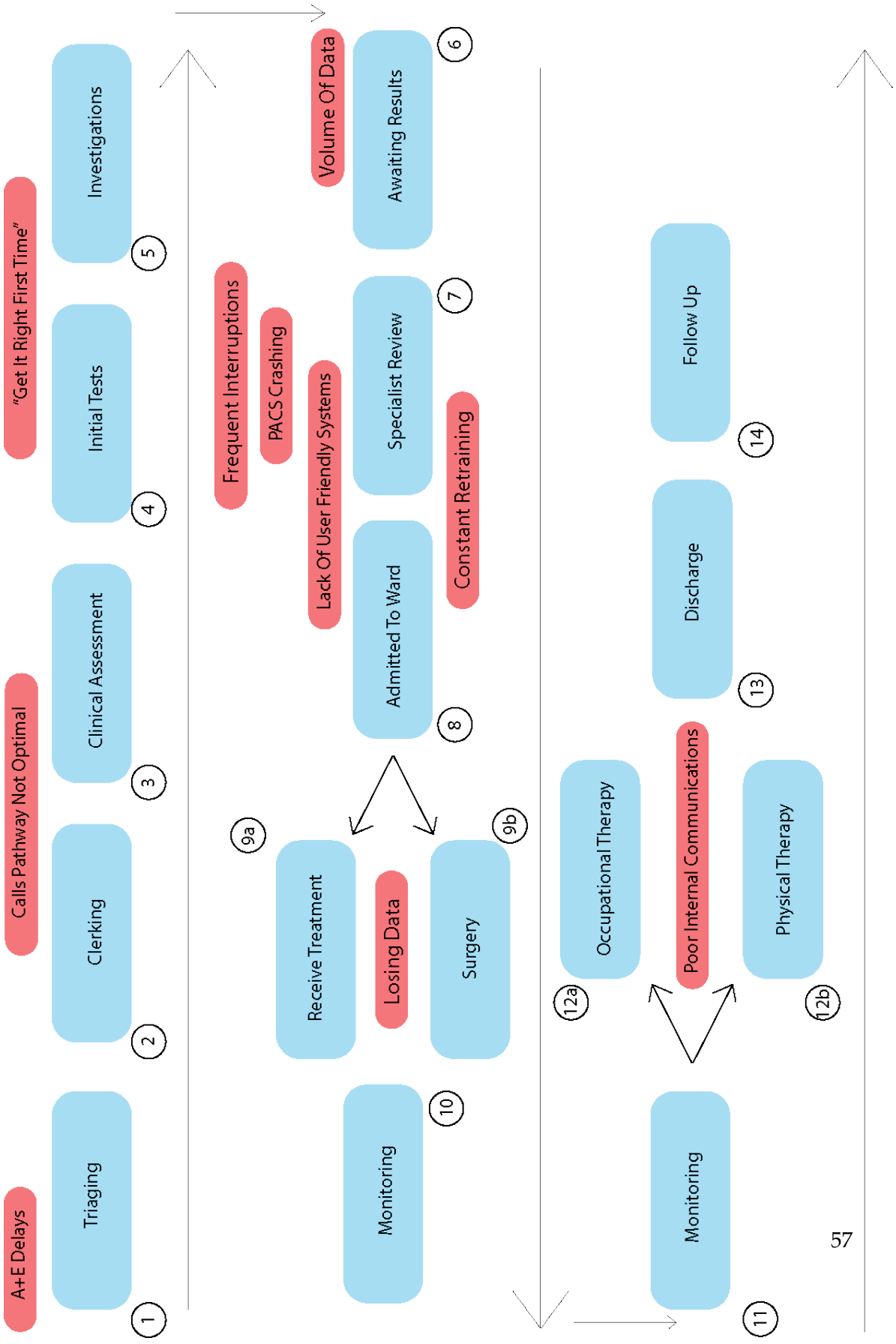


Figure 4.1: Illustration Of A Timeline Agreed Upon By Group 1 To Be "Comprehensive", With Areas Of Friction Highlighted In Red

#### 4.4.1.1 Points Of Failure and Friction

Across the 2 iterations, there were a total of 12 unique points of failure or friction identified – there was considerable overlap between the focus groups, and so here we have presented an example timeline from the first focus group that contains a series of common steps, and have displayed all 12 points of failure from both focus groups along the timeline.

- Accident and Emergency Over Capacity
- Calls Pathway Not Optimised
- Emphasis on "*Get It Right First Time*"
- Ubiquitous Delays [sic]
- High Volume of Patient Data
- Internal Communication Failures
- PACS Crashing
- Loss of Required Patient Data
- Frequent Interruptions when Working on Tasks
- Too Many Individual Systems
- Lack Of User Friendly Systems
- Constant Retraining Required

### 4.4.2 Wish Lists

In total, we ended up with 12 total factors across the two groups that ended up on our *Wish Lists* – in both iterations, participants did not decide to focus their discussions on a singular design idea or point of failure, and instead engaged in free-form discussion about functionalities and factors they believed would universally help their experience. As a result, examples of points of reference for the items on each *Wish List* would be tied to personal experiences or familiarity with 3$^{rd}$ party systems they utilise outside of the workplace or have seen at other Health Boards. This gave us more of an idea of how participants felt existing systems weren't up to their standards than a piece of

speculative design. Figure 4.2 shows an illustration of all of the factors included in the *Wish List* activity across both focus groups, sorted into the categories that will be discussed in the *Discussion* chapter.

- Peer To Peer Communication
- Instant Messaging
- E-Request Functionality
- Accessibility Across Devices
- Integration Across Systems
- Singular Point of Implementation [sic]
- Universal Access To Data
- App Compatibility
- Fast and Stable PACS
- High Quality Equipment
- More Manpower
- Cloud Backup and Support

## 4.5 Discussion

### 4.5.1 Key Themes

Between all exercises and across iterative groups, it becomes clear that there are 3 key themes that arise as fundamental areas of failure, and were accordingly key focal areas for discussion when participants were conducting speculative design activities.

*Communication*

It makes sense that participants spent a considerable amount of time speaking negatively about the connotations of having to wait for results or delays when clerking and onboarding patients, as communication constitutes the majority of their role in the hospital [229, 174] – the necessity for streamlined methods of communicating was seen as one of

the most important elements that could be improved. This can be seen in the inclusion of *Wish List* items such as "Peer To Peer Communication", "Instant Digital Messaging" and "E-Request Functionality".

The lack of codified channels was highlighted as a key pitfall in the timeline of interactions. One participant described a scenario where a patient was provided the participant's internal phone number by the switchboard, because the switchboard operator was unsure of the appropriate channel of communication. Overall, telephone communication was viewed negatively, consistent with existing research on radiology [115]. This was heavily informed by personal experiences of radiologists having to communicate with the relevant clinician:

"*If you as a radiologists that's doing a report find something that actually needs to be urgently relayed . . . you then pick up the phone to ring them and you might spend 45 minutes trying to get hold of them because there is this difficulty with the numbers they might share bleeps or they put down the on-call line or the ward phone*" (P7)

"*If you're doing a report and you don't answer [the phone] because you're at a pivotal part of what you're doing, it'll stop ringing but it'll just ring straight back and keep ringing until you answer it. Sometimes its better if they just come down to the room where you're reporting and just wait for two minutes*" (P9)

As a result, participants frequently referred back to their desires for structured and digitised messaging; participants often drew from their experiences utilising 3rd party applications such as WhatsApp or Microsoft Teams, which were often used to organise breaks or ask informal questions. This aligns with the modern understanding of digital literacy, where users will view uncertain systems through the prism of one they are more familiar with [186, 31], and can help us understand the failings of existing communication systems by comparing them with common and highly accepted implementations in the private sphere.

*Integration*

To properly understand why *integration* permeated much of the discussion of extant issues and how they could be fixed, it is important to understand that NHS procurement

means that software, hardware and all other aspects of diagnosis and treatment have to be identified and procured under strict rules of public policy and suitability. This is to ensure that the correct and most appropriate and safe items are implemented, but has the knock-on effect that, in the holistic process of patient treatment, a series of different vendors may supply each element of the chain. Different brands of hardware, software and peripherals may all be found working together on a singular process.

Here, this was seen as a clear area of friction. Participants opined that interoperability was a chief concern, as often systems did not communicate as well as they were supposed to – having an "*all-in-one*" style of implementation was seen as the pinnacle of design and efficiency;

*"I feel like there's definitely not enough synchronisation between RADIS the reporting system and PACS. So even if I did have a scan of the clinical picture there when I'm doing the reporting, I've got to look to another app to see the actually reason for the report was? It would be nicer If on the reporting system itself I have that information there itself instead of having to go between two different programs"* – P4

This is highlighted in points of frustration such as *too many systems* and *constant retraining* – the high number of digital programs and software that participants were swapping between when completing tasks was seen as a key reason for delays when reporting results and communicating them. These perspectives echo the findings of holistic investigations of procurement and implementation in clinical environments; the vast amount of data being produced means single step adjustments cannot keep up [98, 35]

Similarly, this perceived high number of systems was seen as a major barrier in data storage and access; having multiple programs each independent to a particular health board means that data cannot be shared across sites easily, and participants could not access important personal or patient records remotely or when locuming in a different hospital. This is reflected in the *Wish List* functions of *accessibility across devices* and *universal access* – participants claimed that the variety in programs meant that they had to remember different passwords and logins, each system had data stored in a different place, and layouts were confusing when transferring to a new site. By having a singular point of

implementation, participants hoped to avoid the stress of having to continually remember new systems or risk having their data inaccessible.

*"The more specialist you go, the more logins, and that's a delay because when you first start you job you might not even have all of these or you might have locums that turn up that don't have all of them"* – P5

Participants felt that digitisation was a beneficial aspect of practising medicine compared to having paper based documentation, but saw the fractured nature of un-integrated systems as a hindrance.

*Support and Infrastructure*

Throughout both iterations, it became clear that participants were sceptical that the digital interventions discussed would be as effective as systemic and political change, such as greater funding or a larger workforce. There were several reasons for this - participants were derisive about the amount and quality of equipment available to them, with P5 in particular taking issue with the way digital interventions were being discussed, with statements:

"*Even if you made a really sexy program like Apple, I don't think our computers could handle it"* (P5)

and

"*I worked in A&E in [redacted location] and there were like 3 computers in the fishbowl which is where like the majority of doctors will see patients. . . you have 5 doctors in that room taking turns to use that computer, now there's a delay that can't be solved with a program*" (P5)

demonstrating the belief that introducing novel methods of solving a problem presented would not be as effective as simply procuring more or higher quality computers. We can also see this in the inclusion of terms such as *Manpower* and *Fast and Stable PACS* in the Wish List. *Manpower* specifically appeared to be a rejection in total of the idea behind the activity by participants, making an overt demonstration that they believed a larger workforce would be more effective at tackling their points of failure than any form of digital intervention, whilst *Fast and Stable PACS* is again a statement implying poor quality

of existing infrastructure. The belief that existing systems are not to a sufficient standard is also reflected in the identified points of failure – *PACS crashing*, *Loss of Data* and *Lack of User-Friendly Systems* highlights that this was an area of concern from the outset of the study. This is not an uncommon belief, with Drop et al going so far as to claim that digitisation can even be a hindrance if not carried out with the practitioner's needs placed centrally [105].

The repeated highlighting of ineffective systems seemed to also tie in with the previous connection to existing 3rd party applications and systems, with the tagline

"*We know this exists, why can't we have it in the NHS?*" (P3)

summarising the frustration from participants that they were being asked to discuss how to address underperforming internal systems when they had experience with tools such as WhatsApp that would suit.

The politicisation of design activities when using clinicians as participants has been noted in previous studies [51, 101], with critique of existing hierarchies and working structures frequently dominating discussion – our finding here aligns with this, and highlights how important this factor was to participants; The fact that participants were eager to discuss ways of addressing their identified points of failure without engaging directly with the *digital* aspect of the prompt demonstrated a keen held belief amongst the participant pool that there were ways of solving their problems that didn't require inventing or deploying new ideas, and instead could be examined by taking existing structures and re-evaluating them.

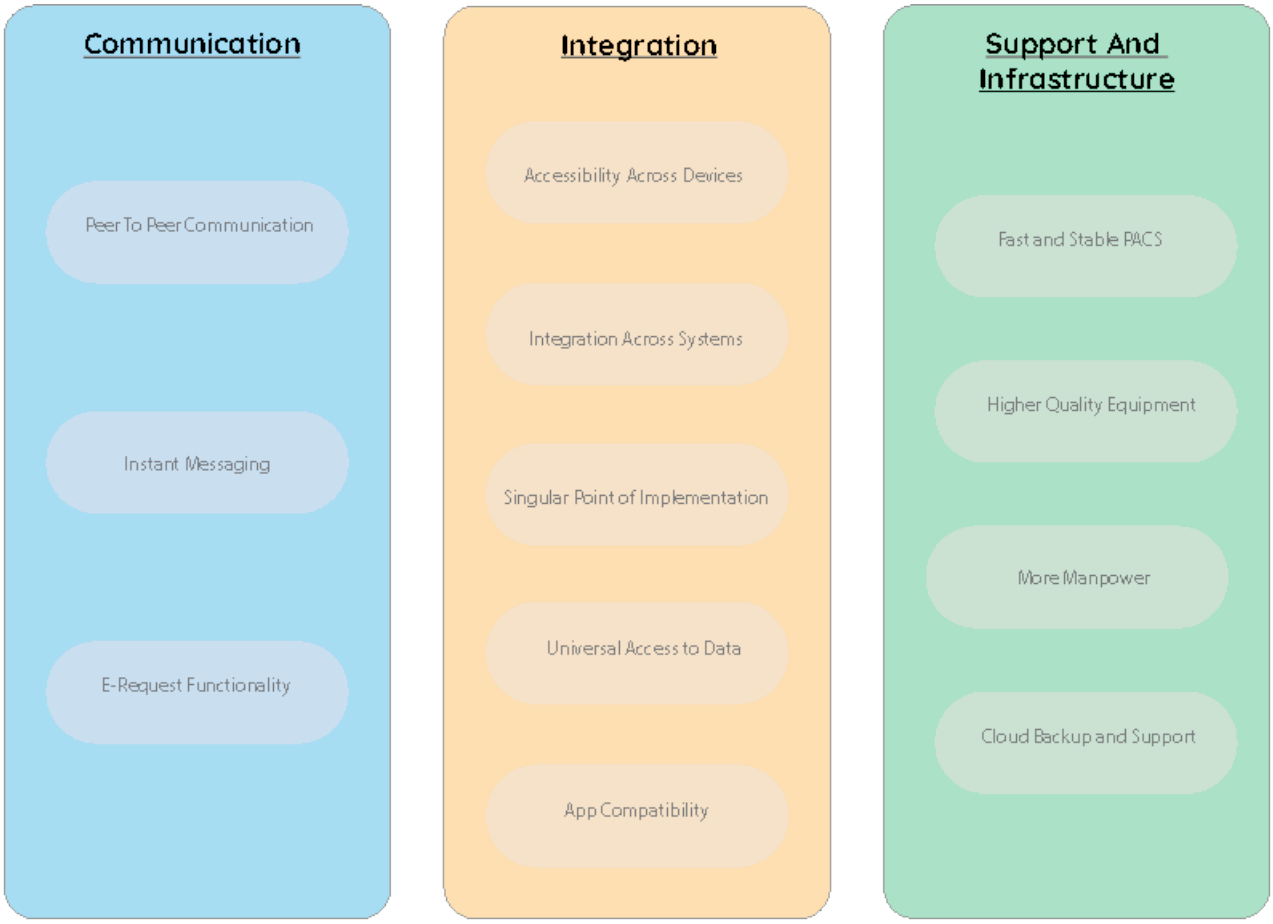


Figure 4.2: Visualisation Of All Items Identified In The Wish List Activity, Sorted and Colour Coded Into General Categories

### 4.5.2 Implications For Further Research

The primary goal of this focus group was to examine areas of failure and friction from the *user's* perspective in order to provide us with an avenue of research. When examining the themes of the study, digital interventions and prototyping do not appear to be a priority for participants; attempts to discuss this in the *wish list* design fiction activity primarily led to more generalised analysis of how the systems currently implemented could be improved, or how better systems did exist they were simply outside of financial constraints.

The latter point proves interesting, considering the framing of this study - due to the

fact that we are solely focusing on NHS based practitioners, the availability and pre-implementation of devices inside the safety critical environment provides us with both a constraint and a focus; we are unable to suggest or recommend to management that they simply replace or upgrade equipment, and designing new tools for implementation would be a futile endeavour due to the fact that they would simply be unable to be implemented in a real-world environment. NHS ethics and regulations would not allow us to do so, and as such we are, in a way, dealing with this financial and regulatory handicap when studying device interactions, and their friction and failure.

This does mean, however, that we have been provided with an avenue of investigation; participants shied away from providing specific tools that failed and how they should be re-designed and instead pointed the researcher towards examining tools that are situated in the safety critical environment and cannot or will not be changed for the foreseeable future, and as such exploration should be refocused around better interactive techniques and methods of teaching and utilising these tools. Accordingly, we determine that work should be done that examines the environment *holistically*, taking into account the organic aspects of community construction that go into how tools are perceived and utilised - we have found in this study that it appears inappropriate to attempt to deconstruct the patient-treatment pathway into individual interactions, and should instead treat the process as a continuation of device interaction and digital intergration.

## 4.6 Summary

This chapter consists of a pair of focus groups conducted with active practitioners to examine their perspectives on the current state of their digital workflow when treating patients. By beginning our research with this high level participant-led construction of the timeline of a patient's treatment, we provide ourselves with a basic level of understanding of the hospital process and construction, as well as the 3 key factors that practitioners believe future research and implementation should address to improve their standards of care.

When examining these 3 key factors, it becomes apparent that participants were concerned that examining stages and points of failure in isolation (much like examining the appropriateness of new systems) was an ineffective way of "seeing the big picture". Participants wanted to examine how the lack of having an integrated singular point of implemented

would *affect* their ability to quickly communicate, and how the lack of good systemic support impacted their ability to be enthusiastic about new digital interventions. Participants wanted to highlight the importance of viewing the patient interaction workflow *holistically* as a whole, as all of these factors concatenated together – this was something they felt was also not seen by management and leadership teams, given the uncodified nature of communication streams and poor usability of systems available to them.

Across our thematic analysis, it becomes clear that

1. Communication is seen as the most important aspect of the patient interaction timeline, coming into discussion when examining points of failure and potential avenues of digital intervention.

2. Participants are not happy with the interoperability and accessibility of the existing systems available to them when diagnosing, communicating and treating patients.

3. There is a strongly held belief among participants that there are ways of addressing existing issues through re-evaluation of holistic workflow, with scepticism shown towards potential digital interventions

Addressing these findings should come in the form of a study that juxtaposes a participant perspective-led approach – the data we have gathered thus far has come entirely from an insider opinion, and as such we should investigate the holistic workflow process from an objective, outsider perspective, to provide a bi-dimensional representation of the situation at hand.

As such, the following study takes the form of a non-interventional observational study of work carried out in-situ. This allows us to examine the accuracy of participant claims whilst investigating and re-evaluating existing behaviour in line with their perspectives on *what should be done*.

# Chapter 5

# Observations Of Radiological Interaction and Communication "In-The-Wild"

## 5.1 Introduction

In the previous chapter, participants frequently referred to issues such as computers crashing, unstable internet connections and individual device failure. They also referred to frequent interruptions due to a lack of standardisation with internal telephone lines and poor decision making from other members of staff. Highlighting these issues to an outsider who they believed could assist in solving them was deemed of significant importance, as there were feeling and perspectives that systems and devices were not designed and implemented with a pragmatic understanding of what conducting radiological analysis actually consisted of. However, as valuable as participant perspectives in focus groups can be (especially when consensus is reached as coherently as it was in our two groups), it should not be taken as accurate without review [256]. There exists the needs for us to "validate" these perspectives, as well as adding another dimension of understanding and analysis from our own perspectives. As such, we wanted to investigate what a radiologist encountered on an average day-to-day basis without potential bias from an insider. This led to the planning and implementation of a non-interventional observation, which we could then assess against the findings of Chapter 4

Here, we decide to investigate the safety critical environment of the radiologist for ourselves, adopting a perspective that will allow us to compare and contrast to the views and points of failure and friction brought up by clinicians previously; this chapter consists of an ethnomethodologically informed ethnography focusing on 11 radiologists and radiographer's interactions with digital systems and devices in-the-wild. We adopt a completely non-interventional "hands-off" approach that allows us to record genuine behaviour in ordinary circumstances, emphasising the mundane aspects that often get lost. By situating ourselves in the radiologists natural working space and observing participants with pen-and-paper notes, we recorded 35 hours of observational data from which we offer a series of reconstructed analyses of how reports are constructed and communicated in an average day to day context. This reconstruction and subsequent analysis presents an environment where the emphasis is placed more heavily on the trust of quality of a practitioner's medical opinions than introducing standardised reports and structured methodologies; methods of communication, written analysis and device interaction are left to the practitioner's preference.

## 5.2 Background

Here, we wanted to focus on authoring and communicating reports, as this was seen as the main crux of what a radiologist was required to do in a given session. Studies on communication have found that whilst practitioners often value "good" communication skills above other personal qualities, they feel that it is under represented as an important tool in the educational phases of their training [54]. This "good" communication is seen as an important part of interacting with both peers and patients – in the former, time is often of the essence and practitioners are required to only communicate necessary information, and in the later complex knowledge has to be adapted in a way that allows non-experts to be informed and confident in the treatment that they are going to be administered [10].

However, in a discipline where communicating results to peers represents a significant amount of the work conducted in a given session such as radiology, there are still significant issues with defining good communication and proper procedure. Communication systems are manually driven, and this allows for a significant amount of human error and procedural variation [20] - up to 55 percent of lawsuits in the United States involving a radiologist name lack of follow up on test results as a factor [212], and there has been

previous study demonstrating that, in an event where the referring clinician is unavailable, radiologists are unsure of the best and most appropriate way of communicating findings to a patient [16].

With this in mind, practitioner uncertainty or negligence is not always the cause of failure in the treatment of patients, and can often be unfairly maligned as the key cause of malpractice – the "blame culture" found in healthcare domains can often mean that investigations are resistant to place device failure or poor design at the centre of a sentinel event [41] Similarly, codified "standards of practice" can often miss the restrictions in place and difficult realities that practitioners face in the real world [18] – additionally, the myriad of ways available to diagnose and communicate findings has had a negative impact on the satisfaction of radiologists [115].

As such, it is clear that there are often issues with communication in medical contexts but it is important to take a full holistic assessment of Work As Done to understand where points of friction and failure exist, and not lay blame at the feet of the practitioner.

## 5.3 Method

Chapters 2 and 3 covered the widespread adoption and implementation of ethnographic methods in healthcare domains, and here we adopt an observational framework based upon works such as Furniss' *Fieldwork For Healthcare chapter HCI Observations On An Oncology Ward* [120], and Coble et al's *Contextual Inquiry: Discovering Physician's True Needs* [83]. We also draw some inspiration from Briedis' previous phenomenological studies of an American radiology head of department – the findings of which called for further and deeper analysis of tool and device use in-situ (which we are obliging here) [57, 58].

### 5.3.1 Ethics

For this study, we completed a full HRA ethics application approved by University Research Governance, NHS Ethics, and each University Health Board's Research and Development department. There were, however, some constraints that affected the nature of our study and the data gathered. We were not allowed to record any form of sensitive information related to patients, meaning no photos or videos were allowed at any point

of the study. As such, we were heavily reliant on pen-and-paper notes. This also meant that we were unable to record our post-session interviews, in the chance that practitioners referenced an event that had occurred during the observation.

### 5.3.2 Approach

We adopted an ethnomethologically informed ethnographic approach for this study, meaning that whilst our output and results resemble a traditional ethnography, we are driven by the why and how of practical observable actions. This takes the form of a "workplace study" (again referred to in Chapter 2) utilised in conjunction with semi-structured interviews with participants, identified by previous researchers such as St Amant, Sellberg and Melconcon as the most appropriate method of capturing both active praxis data, and biographical and tacit knowledge data required to reconstruct the contextual elements of behaviour – we are able to paint a more accurate and appropriate picture if we know more about the participant's background [279, 261, 199].

### 5.3.3 Setting

In contrast to the other work found in this thesis, we found it vital to step out of NIAW campus and situate ourselves in a more naturalised environment – whilst patients are treated at the NIAW, they are often treated by trainees under observation or in collaboration with consultants and trainers. For this study, we chose to conduct our observations in radiology departments of NHS Wales hospitals; this would give us a representation of an average day for participants, in a setting they were comfortable and familiar with. It would also be our best chance to encounter points raised in Chapter 4, such as telephone or peer interruptions and system faults.

We identified 3 sites across 2 University Health Boards (UHBs) – by working across 2 UHBs, we encountered a greater variation in the type of software and systems used, as each UHB was responsible its own procurement and implementation. These sites were The Royal Glamorgan Hospital in Rhondda Cynon Taf and Prince Charles Hospital in Merthyr Tydfil, which were Cwm Taf Morgannwg sites, and University Hospital of Wales in Cardiff, a Cardiff and Vale site.

### 5.3.4 Participants

We recruited 11 participants for this study, 5 of whom were consultant radiologists and the other 6 were reporting radiographers. The difference between a radiologist and a radiographer relates to training and qualifications – a radiologist has a medical degree, and has specialised in diagnostic medicine, whereas a radiographer is a technician and health professional who does not have a medical degree. The difference in practice, therefore, often comes down to the nature of the analysis provided to the referring clinician who has requested the scan; radiologists will utilise their medical training to offer explanations as to why something may have occurred, and will recommend specific treatments, whereas radiographers will be more descriptive in their reports, relying on explaining what the scan shows. For the purposes of this study, our radiologist and radiographer participants shared similar duties of receiving medical studies, analysing them, and producing a radiology report based on their findings to be sent to a referring clinician.

Of our 11 participants, we were able to observe reports being constructed in several mediums across different disciplines. We observed diagnostics of ward patients, paediatric diagnostics of child and infant scans, neuroradiology of brain scans and patient-facing radiology whereby the patient's data was shown to them in-office. This meant that we saw reports constructed studying plain film x-rays, CT, PET and MRI scans, giving us a wide variety of stylistic and analytic data.

### 5.3.5 Interviews

The author individually conducted two interviews with each participant, a structured questionnaire prior to the observational session taking part and a debriefing interview that we utilised to follow up on any unclear elements of behaviour that we had recorded.

The standardised pre-observation interview was designed to elicit biographical and contextual data that would help to add depth to our understanding of practice, in line with Melconcon's principles of Context In Use; we asked 5 questions on demographic and background information, as well as 2 questions related to positional and perspective data that would help to inform more general elements of being a radiologist [199]. These questions would also serve as an introduction to the study as an opportunity for the participant to refamiliarize themselves with the nature of being a subject of observation,

as well as allowing us to reinforce the non-assessed nature of the study – part of radiology training is being observed and assessed on performance, and we took this introduction as an opportunity to dispel any belief that the accuracy of participants findings would play a role in our observation.

1. Do You Understand What This Study Is About?
2. How Long Have You Been Practising In This Field?
3. Where Did You Train?
4. What Modality Of Radiology Are You In?
5. How Do You Write Reports?
6. How Do You Perceive Radiology In Relation To Other Fields Of Healthcare?
7. Do You Have Any Questions Before The Study Commences?

This interview was recorded on a digital handheld recorder to ensure full accuracy of participant's words, with full transcription happening after the session had ended. Basic thematic analysis by the author would then be applied to each transcript, in the form of several read-throughs with open coding before a concatenation of these codes for common themes and discussion points.

The second interview was unstructured, as served as an opportunity for us to follow up on any aspects of behaviour that we felt unqualified to comment on without clarification, or unusual findings that we wanted to discuss with participants. This includes sentinel events that occurred over the session, examples of device failure or fault, and events or actions that seemed to provoke a notable reaction from participants (for better or worse). These interviews were not audibly recorded to avoid sensitive or patient data being recorded, and were instead recorded with handwritten notes that were transcribed after the session had ended to provide "colour commentary" to the observation data.

### 5.3.6 Observation Framework

In order to maintain structure throughout these observations, and to allow for more codified analysis methods, we adopted a series of observational dimensions and methods

laid out by Spradley in the chapter Doing Participant Observation. Following these dimensions gave us a guide of, in Spradley's words, "what to look for" and allowed for better categorisation of notes and observations [277, 278]. The framework comprises of using the following dimensions of Space; Actors; Activities; Objects; Events; Time; Goals and Feelings in a matrix as a way of orienting oneself to the behaviour and asking specific questions that can be used to reconstruct a session [217].

We used this framework in conjunction with the methods laid out in HCI Observations In An Oncology Ward [120] to produce a methodology that focused on being flexible and lighter on structure than traditional ethnomethodology. We embraced the improvisational (and occasionally chaotic) nature of medical practice in our own data gathering, allowing us to take advantage of unplanned and seemingly random events such as interruptions, emergency medical cases, and patients requiring actionable treatment mid-session. We were vocal and overt about embracing this nature with participants, as so to assure them that emergency or unexpected events would not "ruin" a session and that we were content to stop-and-start data gathering as the situation demanded.

### 5.3.7 Session

The observation would be carried out solely by the author alone by themselves in every instance. As described in our Approach chapter, this work was sponsored by the NIAW and as such participants were aware of the collaborative nature of the research - even so, some participants needed to be reminded of the purpose of the study and what it entailed prior to the session beginning, and we made sure to inform participants that there was no assessment of their diagnostic ability taking place.

A typical session would begin with an email conversation discussing when would be best to observe each participant: many participants had irregular work shifts, with multiple duties across different sites, and so would typically not be reporting on studies from 9-5 in an office. Once an appropriate time had been agreed upon, we would meet the participant in the waiting room or foyer nearest to the radiology department, as the reporting rooms and equipment were often behind locked doors. We would conduct the preliminary interview and gain written consent, then encourage participants verbally to "ignore" us; many of the radiologists were experienced educators, and often junior doctors

will sit in on their sessions to learn about the practice of diagnostics. This has the unwanted effect of implicit verbalisation and explanation of their actions, something we specifically asked prior to the observation to discourage - for the best results, the participant should continue as if we were not there. Whilst we would make the best effort to be "unseen" by sitting in the back of the room and attempting to be as silent as possible, there was no way for participants to completely forget my existence, but by pre-empting it with a conversation we limited these verbalisations to a minimum.

Each participant would typically worked in a closed personal office either alone or with one other person. Radiologists had their own offices, and radiographers would work out of unoccupied ones where the resident radiologist was working somewhere else in the hospital or at a different site. The exceptions to this were P8, who spent half of their session working out of a group space full of radiographers and residents, and P9, who was participating in a program that allowed patients to discuss their scans immediately with the radiographer after they were taken. In the session with P9, notes would be made whilst the scan was being analysed, but when the patient entered the office to receive their results we had to halt note-taking and leave the room until they had left. Most sessions would be 3 hours long, as this comprised an average "shift" for an NHS reporter, who's contracts deem 3.75 hours to comprise a session. These sessions were not rigidly timed, as breaks would be taken and participants would often need to leave to address things that had come up across the site. All but one of the observational sessions started at either 9am or 2pm after the lunch break, with one singular session occurring in the evening starting at 6pm.

The participant would then work as normal, authoring reports and addressing interruptions from peers and colleagues. Notes were made on an e-ink tablet as this allowed for low profile writing, drawing and collating of notes, as well as digitisation after the session had ended. Notes would be on anything we deemed to be under one of the Spradley dimensions - this could include the layout of the room and the desk, how the radiologist speaks to their peers, how they interact with their devices, and any distractions, system failures and mistakes that occur. This note making was not exhaustive, but the intended effect was to accurately be able to reconstruct what each "session" comprised of in a thick descriptive manner. More attention was given to capture the nature and content of the

observation than to record everything in a photographic manner.

After a full session had come to a natural close, such as the participant needing to take a break or finishing a large report, we would approach them to indicate that the observational period was over. Care was taken not to interrupt the participant in the middle of a task, as it is important to note that they are working with real patient data and incurring any error may have serious ramifications. After informing the participant that we had finished observing, we would take the opportunity to address actions and events that occurred that we were unsure of for clarification. Discussion with the participant on this topics would also be recorded through note taking as it would be unstructured and ad-hoc, and so there was no guarantee that the content would not record sensitive information. After this discussion had taken place and we were confident we had understood and appropriately recorded the session, we would leave the hospital.

Interviews would be transcribed immediately, and notes would be transcribed to a spreadsheet assigned under the Spradley dimensions as quickly as was practical. This was done to ensure that the session was fresh, as the notes by themselves may have been subject to ambiguous reconstruction if left for too long after observing the participant. This marked the end of the data gathering process for each participant. Clinicians are notoriously hard to "get a hold of", and so in some cases we would record two sessions in a day, or would be working back-to-back days in a single site before not observing another participant for several days. This also pressed the importance of transcribing the data quickly, as so to ensure that we were not quickly left with a large backlog of ethnographic data.

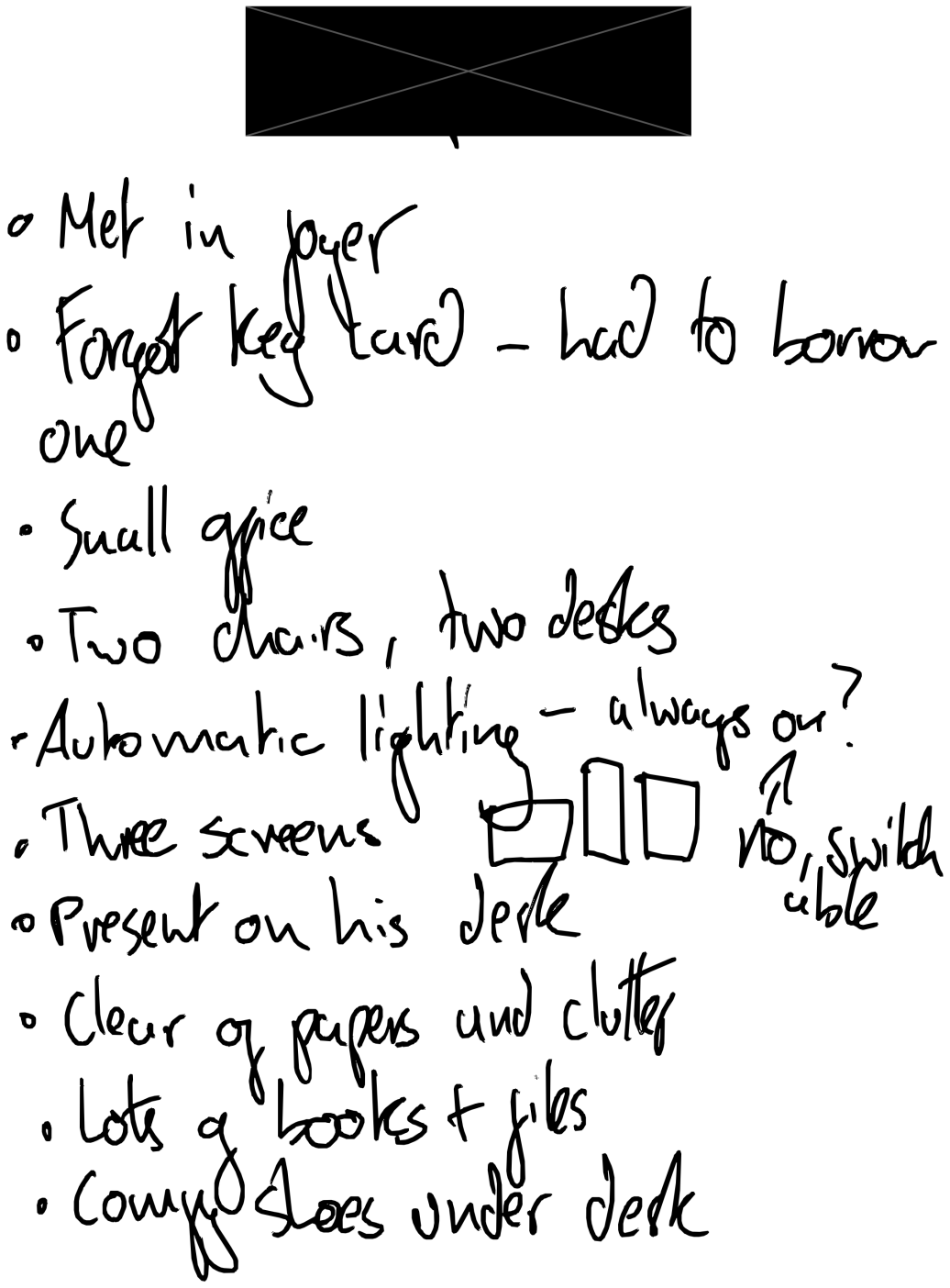


Figure 5.1: Example Fragment of Notes P1's Session, Showing Elements Of "Space" Dimension

## 5.4 Analysis

### 5.4.1 Interviews

The author would conduct two passes of analysis on pre-session interviews. The first pass was to extract demographic data, and the second pass was to conduct more in-depth thematic analysis on our contextual questions relating to perspectives and positions of participants. We adopted a lightweight method of thematic analysis based upon Braun and Clark's methodology, comprising of coding each transcript individually before reading all transcript codes together to determine themes [55].

### 5.4.2 Observation

Whilst organising participants data structurally according to our Spradley matrix of dimensions was an appropriate method for gathering, we adopted a different approach for analysis. We adopt the analytic method set out by Crabtree (developed by Geertz 1973, in turn based upon Ryle 1968) of "thick descriptions" being formed from reconstructions of our data [86]. This comprises of reassembling each session from notes into a description of behaviours that encompasses the process of "writing a report", from which we can extract unusual or notable examples of interaction for discussion. We add contextual information and combine different dimensions together to write an in-depth description that paints a picture of the whole process. The result of this analysis is a series of vignettes that serve to demonstrate common and uncommon elements of interaction and communication that take place during a typical reporting session for a participant.

## 5.5 Results

### 5.5.1 Interviews

Our demographic questions were useful in illustrating the diverse range of participants, with considerable variation in speciality and number of practising years. However, when it relates to the interactive praxis that we will be examining later in this chapter, it is worth highlighting an apparent lack of relationship between where participants trained (almost all in the UK, with P1 spending some time practising in Australia), their speciality or their experience in the field and their interactional behaviours and patterns. We noticed some differences between the behaviours of radiologists and radiographers that shall be

| P | Occupation | Speciality | Years Spent Reporting |
|---|---|---|---|
| 1 | Radiologist | MSK | 16 |
| 2 | Radiographer | Chest and Abdomen | 2 |
| 3 | Radiographer | CT and Plain Film | 6 |
| 4 | Radiographer | MSK | 4 |
| 5 | Radiographer | Chest and Abdomen | 3 |
| 6 | Radiologist | PET and CT | 6 |
| 7 | Radiologist | Paediatrics | 23 |
| 8 | Radiologist | Paediatrics | 5 |
| 9 | Radiographer | Paediatrics | 9 |
| 10 | Radiographer | MSK, Chest and Abdomen | 4 |
| 11 | Radiologist | Neuroradiology | 2 |

Table 5.1: Observational Study Participant Information

commented on, but overall there was such significant personalisation across the participant pool that is makes it difficult to track any connection between our demographic questions and any possible trends.

*How Do You Write Reports?*

We found that participants were hesitant to give a singular, all-encompassing answer to how they write reports. There were several reasons for this – different types of scans required varying lengths of reports, and various body parts may require different ways of going about analysis (P4). The "report" was not viewed as a static, singular object, but as a fluid, case-by-case document.

Of all 11 participants, only P1 offers a decisive and cohesive answer for the structure of their reports: "title, clinical indication, report content, and a conclusion if it's a long one". Instead, the majority of participants qualified their answers with the caveat that there was a strong element of personal choice and preference for how their reports were constructed. Comments such as *"I guess I do, in my head"* (P9); *"I would say I do, but it's hard to describe"* (P2); *"There are things I aim for"* (P8); "*If you asked for my style, I would say 'succinct' "* (P7) indicate that, whilst participants may have their own way of structuring reports, they are aware that it is not a standardised practice and there is significant variation between practitioners and across health boards – this was noted by P7, who claimed "I read lots

of other people's reports, and just skip to the conclusion" showing both the variation between reports and strategies to mitigate it.

P6 also noted *"There have been studies that show reporting proformas have value in their consistency"* drawing attention to the lack of structure found across reporting procedure, demonstrating again the keen awareness that these diagnostic practitioners had that there is an overt lack of standardisation, but that it is accepted as the norm regardless of efficacy.

*How Do You Perceive Radiology In Relation To Other Fields Of Healthcare?*

When asked how they perceive radiology in relation to other fields of healthcare, participants overwhelmingly offered two particular aspects of radiology that make it unique: other fields of clinical medicine's reliance on diagnostic imaging, and the lack of patient interaction. Due to the prevalence of imaging in the patient's journey, participants perceived themselves to be one of the most important stops in the hospital; *"One of the specialities all others rely on"* (P11); *"There's a lot of other disciplines that wouldn't be able to function without radiology in the background"* (P4). An extension of this belief seems to be the perspective that radiology is misunderstood by other specialities. Participants discussed that many other clinicians visit the radiology department to get a clinical opinion expect their request for referral to be immediately approved and the result to align with their suspected outcome; *"I don't think a lot of people understand radiology from other disciplines"* (P6); *"Their perception doesn't always line up with your perception, and that's why you're important"* (P9). These views compound to produce a self-reflective perspective of the radiology practitioner to be (in the words of P3) a ***"Gatekeeper"*** of diagnostic imaging.

Regarding the lack of patient interaction, participants deemed this to be one of the most important elements that sets radiology and radiography apart from other fields of clinical medicine; *"I think there's generally less patient interaction"* (P9); *"I don't have a lot of patient contact at all"* (P11). Practitioners self-identified that this is unusual, with P7 claiming the ability to be comfortable with it to be one of the key skills that a radiologist should work on; *"You have to have the ability to focus and have your attention on something for longer"*. This connection to patient data but not the patient themselves was also established as a separation factor between radiology and other disciplines; *"We've a very different*

*relationship to patients"* (P10); *"You have to bear in mind that we've not seen the patient when we report on them"* (P3). One participant claimed: *"I don't want to use the term 'conveyor belt', but we'll only see a patient's data for a very short amount of time"* (P10). These two aspects combined show that radiologists and radiographers view their work as very different to how traditional doctors and practitioners may work in the same hospital, treating the same patients, and perceptions of them should be adjusted as such.

### 5.5.2 Observations

In total, we gathered 35 hours of observational data across 12 sessions – P1 asked us to conduct a 2nd session at a different site, as their 1st session took place at a very quiet evening and they requested we re-visit them at a busier time for an additional perspective on their roles and duties. Here, we will first outline a general "overview of practice" that synthesises the common methods and behaviours that constitute writing a report, followed by dedicated investigations into different aspects of interaction and communication. It is worth addressing that these thick descriptions come from our own observations, and may not align with the "textbook" methods outlined in training materials.

### 5.5.3 Writing A Report

Most participants conducted their work from a private office, comprised of a desk with a desktop computer, keyboard, mouse, computer monitors (ranging from 2 to 4, often oriented in portrait layout for a better perspective of the medical study), a dictaphone microphone, and assorted paperwork. The private office is in a secluded department away from the wards, with the office sections inaccessible without a key or RFID access card – this results in a quiet, calm environment populated by consultants, nurses, secretaries and registrars only. The offices themselves are assigned to each consultant, with personal effects and items that ensure a level of personal comfort - for example, as shown in fig 5.1, P1 had spare shoes under their desk to allow them to walk around in more comfort than their current formal attire. In the case of P8, their work was conducted in a shared communal space filled with registrars and consultants all conducting analysis on their own assigned medical studies at once, resulting in a slighter louder but still quiet environment, with conversations happening cross-desk but the volume of speech still kept low.

The office is kept dark to allow for better viewing of each screen, all of which have

their brightness turned up as far as the monitor will allow. The reporter positions themselves at their desk and finds a comfortable position in their chair that allows them to relax and very remain still for minutes at a time. Participants 1, 4 and 10 had music playing in headphones whilst they worked – P1 wore an open-ear headset connected to a CD player, whilst P4 and P10 wore one side of a pair of earphones connected to their mobile phone. The room, barring interruption, was silent bar the sounds of air conditioning or keyboard and mouse noise.

After finding a comfortable position and completing auxiliary tasks such as answering emails and messages or clearing unwanted applications from the monitors, the reporter turns their focus towards the task at hand. They select a study from the PACS, shown in a long list of names and study details such as the medium, basic patient information, and when the scan was taken. These studies have been assigned to them specifically, and whilst they can pick and choose which scan they want to report on, the onus is on them to complete as many as they can in a session.

Having selected a study to address, they will be presented will a large collection of information. This includes the scan itself, whether a static image of an x-ray or a whole slide deck of slices that makes up a CT/MRI, the Request for Referral Document, and medical information on the patient-at-hand. The initial step is to examine the request for referral, as this contains information on the nature of the patient's illness or injury, the reason for the scan, and the result the clinician is hopeful for. These requests are often handwritten and digitised through a flatbed scanner, meaning there are occasions where the reporter will be unable to read the referrer's handwriting, and this will be noted in the end report. After reviewing the request, the reporter will examine the patients' medical information to ensure that they are informed and aware of what they will be viewing, giving potential foreshadowing to incidental findings, and also allowing them to tailor their treatment advice based on the patient's medical history. The screen positioned on the furthest left holds an empty text box, where the report will be authored by the reporter speaking aloud. After ensuring that it is empty, the reporting begins.

They trigger the speech to text software to begin recording with the press of a button on the handheld dictaphone, with the textbox corresponding in kind by changing

the background of the box to green, signifying the recording is on. P7 also implemented an additional feature where the reporting window would chime when recording was triggered on or off (the chime being the same sound in both instances, leading to some confusion later down the lien). There was a further indication that recording was taking place, with a small microphone icon at the bottom of the reporting window showing a volume meter, indicating to participants the volume of input into the microphone.

As they begin reporting, they vocalise descriptive aspects of the study, including details based upon the patient's history and (in the case of radiologists) what treatment they believe should be administered. As they are speaking, their words appear at slight delay in the reporting window, with a pause being needed for the software to process long sentences. Tools to manipulate the scan are selected from a toolbar in the PACS window, such as zooming in or out, adjusting contrast or brightness, and using a cursor operated ruler to measure the size of organs, fractures or tumours. After the initial assessment is completed, further passes over the study are conducted to check for incidental or immediately actionable findings.

Whilst the report is being authored, there are continuous checks over towards the text window to ensure that their words are being recorded, and to perform small proofs to eliminate major errors such as missing words or glaring spelling mistakes. A full proof is conducted after the report is finished. We will address proofing methods and common errors found in Chapter X. Once the reporter is satisfied that the report is complete and accurate, they sign it and export it to the Radiology Information System for the referring clinician to read, and the process begins again.

### 5.5.4 Speech Recognition and Interaction

*Focus*

As alluded to above, focus is often split between the medical study in front of them and their words being transcribed in real time in an adjacent window. Their gaze moves across all of the monitors frequently throughout the reporting process even when posture remains still, and it is clear that they are keeping some element of attention kept on the text as they will pause in their speaking when there is a significant delay in their words appearing on screen.

This delay in waiting for the software to "catch up" was a particular point of friction for participants, with P7 calling it "daft" and claiming the software will often mistranscribe important terms and sentences in a report; "If it's going to leave a word out, it'll always be the most important one", with an example happening in that session being "in changed" recorded instead of "unchanged", demonstrating the prevalence of mistakes that happen on behalf of the software, as opposed to the user.

P7 also implemented the aforementioned auditory feedback function in their PACS in the form of a chime when the recording was triggered on and off. They were the only participant to implement this function, but it caused significant confusion on several occasions – due to the sound of the notification being the same for recording being triggered on as off, P7 was often left unsure as to what state their report was in. On 2 separate occasions, they began vocalising their opinions only to realise that the reporting window was not recording. Even without the auditory function, this happened with several other participants; P6 began speaking without having properly triggered the recording function, and P5 accidentally left the report open, recording some incidental room noise and personal musings that were then transcribed unintentionally.

*Vocalisation*

Another interesting aspect of interaction between the reporter and their reporting software was the volume and timbre adopted when communicating with the handheld microphone dictaphone. This dictaphone was a ubiquitous device found in every reporting office, with a grille at the end of the device signifying where the microphone input was – this grille would be brought as close to the reporter's mouth as possible, and then vocalisations would be made quietly, in a hushed tone with little intonational changes resulting in a monotonous, unusual speaking voice. This would only occur when the reporter was actively authoring reports, as when a colleague would enter the room or the phone would ring, a code-switching like change would be observable, with the reporter's voice becoming louder with natural intonation resuming.

This appeared to be a conscious decision, and resembled traditional code-switching found in other language environments despite it being with a digital system as opposed

to another human being. It does, however, contrast with existing understandings of interactions with digital Voice User Interfaces – literature suggests that participants hyper-articulate and speak louder with simplified terms to ensure proper recognition occurs [214, 213, 228, 306, 283, 147].

When we addressed this with participants, only P1 and P2 were completely unaware of their voice changing when they started reporting. Other participants were aware that they utilised some form of code switching, but the reasons given different and were often only backed with anecdotal evidence and folk theories. P5 claimed *"It works better when I do it like that"*, and P3 claimed *"The dictaphone doesn't like noise... so you try and get it really close to you"*. No participants said they had read metrics to influence their choices, or had been instructed to use their dictaphones in this manner. It should be noted that P1 also had loud music playing through a headset, P8 was working in a communal setting, and P6 would occasionally hold the dictaphone away from their face to get closer to their monitor and continue speaking, but none experienced a high level of mistranscription errors.

*Internal and External Communication*

When communicating results, the primary method was utilising the PACS software to export the finished report to the referring clinician. However, this was only used in a situation that required no clarification, additional details, or actionable findings. In the case that something more than the finished text report was needed, reporters had a variety of methods at their disposal. We witnessed at least 5 methods other than exporting a finished report to communicate with a peer or a patient, through a wide array of mediums. These included email (Outlook), instant messaging (Teams, WhatsApp, SMS, PACS internal), telephones (Internal landlines and personal mobiles), writing pen-and-paper notes or Post-Its for colleagues to read upon their return or to leave in places where they would be found, and finally leaving the office to discuss something in person either in the department or on the wards.

Most notably about this variety of communication methods is that there was no codified reason for choosing how to communicate a result, and it was instead dependent on 3 factors – the nature of the issue; the reporter's personal preference; the method most likely to get the attention of the receiver. For example, the radiographer participants would

communicate with each other via WhatsApp and SMS, but would attempt to get in touch with more senior members of staff through email or on the internal landline telephones; this was because radiographers knew that they were more likely to get an instant response using each medium depending on the receiver, and that it was more appropriate to discuss a serious question over the phone to allow for better question-and-answer. P9 claimed that, whilst the internal PACS messaging was designed to be a catchall method of communicating in-house, "it's not very good, we'd rather just try and chat through WhatsApp or text if we're trying to find each other".

This lack of structure and codification was evident in the case of actionable findings. In a key example, P6 found an emergency in a scan that was of high priority, and needed to be acted on immediately. They attempted to get in touch with the referring clinician over the telephone by calling the landline switchboard of the relevant health board. The wait time to be connected to the receptionist took over 5 minutes, and since P6 wanted to describe the actionable finding in detail they did not dismiss the scan, and so spent the whole waiting time unable to do any other reporting. Upon being connected, P6 was told that the switchboard was unable to forward the call to the relevant physician (we were unable to ask the reason in this case), and as such P6 used a search engine to identify the physician and their contact details, called them with their personal phone, and left a voicemail requesting a callback to discuss the emergency. This process took over 10 minutes, and was evidently a victim of having no codified method of the reporter getting in touch with the requesting clinician.

*Interrupted Activity*

A final point of notice that made up a standard day in the life for the reporter was the frequency and disruptive nature of interruptions. Here, we are defining "interruption" as anything that resulted in the reporter having to pause their current task to address. Interruptions were a ubiquitous aspect of constructing and communicating reports, to the extent that they were not seen as a serious hindrance in the eyes of the reporters themselves – *"Radiologists are used to working in an environment like this, with lots of interruptions"* (P7).

The nature and severity of these interruptions varied. The most common form of interruption was a peer or colleague entering the private office space to speak to the

reporter; this could be something trivial and social, such as a weekend plans or café recommendations, or a serious matter that needed quick and decisive action such as help with a scan or a high priority case in the hospital. Since the majority of participants kept their doors closed, the process would follow social norms of the interrupter knocking on the door, waiting for the reporter to respond, then entering and proceeding with the query. This requires a total shutdown of the task the reporter was completing, but allows for the reporter to respond if and when they are ready, affording them the ability to reject an interruption if it was of serious importance. The same goes for another common form of interruption in the form of an intrusive notification such as an email or messaging notification that causes a pop-up on the monitor – whilst in most cases it requires the medical study to be put on hold, the reporter has the flexibility to mentally save their place, finish a thought, or leave a pointer for when the task is resumed.

This was seen as adding time to an already speed-focused series of tasks, but the cognitive disruption that it proposed was not especially concerning to participants. In the case of in-person interruptions, participants would often implement a physical barrier by facing away from their monitors and placing their dictaphone on the desk to allow them to properly focus on the conversation. Once the interruption was finished, it would require re-reading of the report, often erasing the last line and giving themselves a "run-in" to continue and finish it. On the other hand, phone calls were seen as particularly disrupting, as participants were trained to address the call immediately - *"We're told to answer the phone as quickly as we can because sometimes it can be very important, but we don't know until we've picked it up. That can be annoying when its ringing at times"* (P2).

This highlights a key issue with most interruptions that participants faced, that there was the potential that they would need immediate response, but no guarantee, meaning they required participants to halt their activity immediately to address them. In the case of P2, we counted at least 20 interruptions from our observational notes, all of which required the report to be abandoned, the interruption addressed, then the report continued or restarted. This results in a significant amount of refocusing, and the implication that the reporter cannot wholly focus on the medical study for fear that they will soon be interrupted.

## 5.6 Discussion

### 5.6.1 Structure and Standardisation

These vignettes of interaction and practice have demonstrated the inherent lack of standardisation that permeates the practical of being a radiologist. Whilst there is still a considerable amount of regulation surrounding the general care administered to patients, there was a high amount of personalisation and idiosyncrasy in the execution of day-to-day tasks. Participants were aware and acknowledged this factor, explaining that reporting styles evolved over time and is unique to each practitioner – they were, however, undecided on if this personalisation of medical information had a significant effect on how well it was communicated to the referring clinician.

Efforts to formally standardise reports have been previously made, with high levels of satisfaction from referring clinicians (i.e. those receiving the reports), but were widely disliked by the radiologists themselves [123, 157] - further study has shown that due to ways in which radiologists train, such as by observing senior members of the community and advice from peers, it can make it difficult to adopt new elements to their workflow such as standardised structure to reports [60]. The implication here is that if standardisation of interaction and praxis is to be widely adopted, it must begin at the earliest levels of education and become entrenched before it is accepted across the discipline.

### 5.6.2 Cultural Separation

One particular aspect that arose between both self-identification and observation was the cultural separation between diagnostic medicine and administrative medicine; not only are there physical barriers between the departments that inhibit peer to peer communication, but there seems to be a radiological culture surrounding the nature of constructing reports and communicating opinions and results to non-radiologists. Studies have shown that the physical and geographical separation of radiologists can have tangible effects on their mental health and self-perception, resulting in higher levels of attrition and burnout [154, 66], but also that a lack of mutual language and understanding between patient-facing and non-patient facing disciplines can result in poor communication of needs and results [269].

Here, we see that terms such as "gatekeeper" were identified by participants to describe how they felt in relation to other fields of healthcare, and radiologists have their own lexicon for diagnostic imaging that mean they are distinct from other disciplines, affirming these previous studies. For our purposes, this reinforces the idea that interactive systems and tools for communicating findings should take into account that there is an element of intercultural communication happening that does not exist across other medical borders [280].

### 5.6.3 Interaction "Best Practice"

The ways that reporters interacted with voice-user interfaces was of particular note – the fact that participants were aware of their particular vocal techniques and cadences, but were not confident in where they had picked this up from nor whether or not it affected accuracy of how the speech recognition software transcribed them demonstrated the prevalence of "folk theories" [125] in digital interaction - though in HCI this term is often applied to algorithmic bias [323], we can clearly see that mental models of best practice are formed in the *doing* of interacting with these digital devices. Dictaphones and PACS software are subject to rigorous testing, and it is likely that these metrics are available to procurement and implementation leads, but it is unclear whether or not these methods of interaction are rooted in this data.

As such, this makes recommendations for best practice or future design more difficult - not only have we demonstrated a distinct lack of codified technique that requires generic design to accommodate for personalisation, but we also have a lack of clarity in *where* these techniques arose, as influences could come from personal beliefs, advice from peers, or at some point in the training process. Further investigation of VUI-based interaction must be conducted with greater input from the actual user group to understand the specific providence of these techniques.

### 5.6.4 Comparisons To Participant Led Focus Group

It is worth comparing our findings from non-interventional observations of Work as Done to the self-reported issues of participants in our focus groups. We did not identify any cases across our 35 hours of PACS crashing, nor did we experience a serious loss of data from a technical error, which were both identified as serious issues that plagued the clinical workflow. We did, however, witness frequent interruptions, and our case of

P6's actionable findings could align with poor internal communications. It is interesting to note that we can confirm the human-factors issues, but not the digital systems factors.

This goes some way to validating the issues highlighted by participants that hinder their ability to provide effective treatment, but also go some ways to dispel the poor quality and function of digital systems. It is possible that we were lucky not to witness a serious outage during our observational sessions, and equally possible that these issues may be infrequent, but when they occur cause such disruption that they remain the priority to be addressed by those who experience them.

It is also worth noting that a key aspect of the Wish List developed by participants was communication, specifically clearer lines of communication and an all-in-one system. This would align with our own findings that communicating results was an unstructured and uncodified practice.

## 5.7 Summary

In order to address the research question *"How Do Radiologists Interact with Devices In-The-Wild When Constructing Diagnostic Reports?"* we decided to undertake a study that would present us with a holistic and comprehensive image of the radiologist in their genuine working environment.

This chapter consists of an *Ethnographically Informed Ethnography* in the form of a non-interventional observational study, examining genuine patterns of communication and interaction in-the-wild. We observed radiologists and radiographers constructing and communicating diagnostic reports, and from these observations we offer an HCI-focused reconstruction of the radiologist "*at work*". This provides us with a wide and overarching answer to the research question provided above.

A clear next step that has been highlighted is further investigation into the details of constructing reports. We have provided a series of vignettes that serve to illustrate the shape of practical work and structured behaviour when analysing medical studies and writing radiological reports, but this ethnography suffered constraints of ethical approval that meant we were reliant on note taking and reconstructive analysis. Video and photo

data of reporters carrying out tasks would add a significant depth and dimension to the data already gathered, and would serve to answer several points that arose in our discussion section.

We have identified that reports are personalised and unique to each participant, and that the way they are authored involves heavy interaction with Voice User Interfaces. In the next chapter, we will conduct a more codified examination of the ways in which practitioners interact with and collaborate with voice user interfaces when constructing reports.

# Chapter 6

# Exploring Interactions With Voice User Interfaces When Authoring and Editing Reports

## 6.1 Introduction

Chapter 5's observational study identified several ways in which radiologist interact with their voice user interfaces that diverged from our expectations. Specifically, radiologists adopted a particular method of speech that we found interesting – it was low in their speaking register, with long phrases that seemed unnatural compared to normal speech and a monotone timbre. As we established, this method, employed because participants believed "it works better when I do that", juxtaposes existing literature about methods such as hyper-articulation and repetition when people are faced with a VUI that they fear may not be accurately transcribing their words [214, 213, 215].

This was a point of interest in our observations, but due to methodological restrictions it was difficult to accurately examine and codify them. Here, we focus in on this specific aspect of the radiologist's practice – if the diagnosis and medical opinion is the key product of the radiologist, and reporting via Voice User Interface is the nationally recognised and most widely adopted method, then this interaction constitutes an extremely important relationship. We observed that the dictaphone and transcription software is the primary method of recording and communicating the radiologist's primary output, and

as such we are studying it in further detail.

To address previous problems, however, it is necessary to take away an element of the real world by implementing this study in a controlled, laboratory environment on NIAW campus. We augment the lack of in-situ accuracy with the addition of "colour commentary" collaborative analysis from participants themselves. This allows us to couple this study with the wholly non-interventional chapter 5 to provide an all-encompassing understanding of the reporting process in both the wild and in vivid recorded detail.

This chapter presents an ethnomethodologically informed ethnography that examines the relationship between the radiologist and the speech-to-text systems. We specifically focus on technique and its impact on errors, proofing and overall satisfaction and accuracy of end reports. In contrast to the previous chapter, we adopted a controlled approach by utilising a laboratory setting with a facsimile reporting system and pre-cleaned data for participants to report on. This enabled us to record participants on video, presenting us with further options for analysis. We recruited 10 radiologists comprising of trainees and consultants, and allowed ourselves a more in-depth examination of the interactive practices that constitute dictating and editing a report.

We also adopted a collaborative analysis approach to our examination of the data, whereby we invited a sample of the participant pool to examine their own video data and offer analytic and contextual insight into the origins, purpose and efficacy of their practices while reporting. When reconstructed, we offer a representation of a radiologist who has been transformed into a literary agent responsible for the proofing and editing of a free-text block without the appropriate training, operating on a system that offers little in the form of radiologically focused assistance. We offer ways to improve this relationship with the goal of improving end-patient care.

## 6.2 Method

### 6.2.1 Pilot Study

Prior to undertaking this research, we conducted a brief pilot study into basic microphone interaction when analysing images. The main purpose of this pilot study was to examine

and test equipment, but also gave us some insight into how untrained users interacted with dictation equipment when presented with an image to describe.

We created a mock reporting office on Swansea University campus, and recruited 5 postgraduate Computer Science students who worked in the Computational Foundry building, all completing PhD work. None of these participants worked in a diagnostic medicine environment, and all claimed to be unfamiliar with Speech To Text software. The mock reporting office comprised of a hidden desktop computer, a computer monitor, a mouse and keyboard, and a microphone attached to a goose neck stand that could be manipulated by the participant, and a single camera angle facing the participant at 45 degrees to catch their interactions and a view of the computer monitor. The computer monitor displayed 2 windows, one being Microsoft Word Processor and the other a slide show of black and white stock images. These images were procured from the University of Stirling's psychology department, and are intended to be generic images for use in psychological and sociological research [227].

Our microphone would be connected to Word, allowing participants to dictate handsfree to their "report", but we also informed participants that the mouse and keyboard were operational. Participants would be shown an image and simply asked to describe what they saw in detail, with the transcription being triggered by the researcher. When participants were finished, they would ask for the transcription to be turned off, and then examine their transcribed report before confirming to the researcher that they were confident in the report's accuracy compared to their vocalised descriptions. Once the participant had examined 3 images, we ended the session. In the discussion of this study, we will refer to participants as "S" to signify "student" and differentiate them from the participants in the full study.

Analysis for this pilot study was brief, and mostly used as illustrative – footage was reviewed in full with the transcript alongside the playback, and rough notes were made on how each participant behaved. These notes were then compared, and basic themes were decided upon. The transcripts were also reviewed for mistakes, and unintended words being left in.

When reviewing the footage, participants varied heavily in the amount of words they used to describe an image. S5 used an average of 28 words per image, whereas S1 used an average 112, with the average across the 5 participants being 71. There was, however, clear consistency in the way in which participants interacted with the devices when reviewing, proofing and editing their "reports": none of our 5 participants edited with vocalisation, all instead utilising the provided keyboard and mouse to identify and correct mistakes. No participants touched the microphone provided to move it further or closer away, instead choosing to position themselves in a way that was comfortable but still close enough for the audio to be picked up.

Additionally, for all participants but S3, proofing would consist of reading the report aloud, often accompanied by following through the report with the mouse cursor as the report was read. All participants barring S1 left at least 1 count of "non-analytic speech" (i.e. talking to themselves or the researcher) in their reports, and all participants had at least 1 unfixed error across their 3 image reports. Proofreading itself rarely lasted more than a few seconds, consisting of a brief re-read of the report aloud before moving on, often missing these mistakes.

The purpose of this pilot study was, as described, to test our equipment, but also revealed some basic insights into how untrained users interacted with the mock VUI setup: microphone interaction was minimal, with participants overwhelmingly choosing to immediately use a keyboard and mouse to interact with the computer; proofreading was brief, and often inaccurate, but this study does orient us to how an entry level user, unfamiliar with these methods of reporting, behaves when presented with this scenario.

### 6.2.2 Approach

As utilised in our previous study, we held a particular emphasis on ethnomethodological findings – we wanted to examine how mundane behaviours came together to form a holistic representation of interaction praxis. Again, our aim is not to rigidly codify all instances and methods of interaction, but instead to illustrate tacit "common sense" behaviours for deeper analysis and extrapolation. In addition to this method, we also employed a "fact checking" method of collaborative analysis. As computational researchers we had reached a level of vulgar competency that allowed us to design

and carry out this study, but when it came to analysis and discussion we wanted to avoid emphasising aspects of behaviour that were unique to the laboratory environment. Similarly, we wanted to ensure that our findings were properly informed and generalisable.

To do this, we involved a sample of participants in our analysis as collaborative researchers. Collaborative ethnography as a method is often used when researching vulnerable groups that the research team feels are far removed from their own experiences and mindset, and so we felt it was appropriate here to employ when studying such specialist users. We invited participants to comment on their own interaction footage, and offer insight into the reasons behind various behaviours and frequency of them in real-world practice. This collaborative analysis took place alongside and after initial viewing discussion, and so acted as supplementary "colour-commentary", adding contextual information and weight to findings and conclusions.

### 6.2.3 Participants

For this study, we again chose participants who were active practitioners within NHS Wales, specifically at the Cwm Taf Morgannwyg and Cardiff and Vale Health Boards. We did not elect to use any radiographers, as we wanted to focus on a particular subset of the diagnostic population who have all shared the same training routine – there are a variety of ways one can train to become a radiographer both academically and through vocational practice, but radiologists have a set syllabus of teaching in order to qualify. This meant that, at least for most of our participants, they learned how to use VUI systems in the same way. For the same reason, we drew from a participant pool of both consultants and trainees at the National Imaging Academy to examine potential differences in interactive technique between those who had been using the VUI systems for years and had experience reporting in radiology prior to their implementation, and those who have learned to report with VUIs and were relatively new to the practice.

The trainees we selected from this study were experienced enough to report on real patients unsupervised, and so we did not foresee issues with their abilities to examine medical studies. Of our participant pool of 10, the split was even with 5 consultants and 5 trainees. Participants had a range of experience, with the least experienced having worked as a reporter for 3 years and the most having more than 25 years. 10 participants took

part in the lab study, with 4 participants returning to collaboratively analyse their data. 3 of those who participated in the collaborative analysis were Consultants. Difficulties with active clinical practitioners as participants in research has been addressed frequently, and this was something that was taken into account when recruiting and conducting for our study [120, 242]. Furthermore, we refer to the ethnographic principle that human activities contain their own means of generalisation, and as such attach significance to our findings from these sessions [86, 88].

| Participant | Role | Plain Films Reported | CT Scans Reported | Took Part in Analysis |
|---|---|---|---|---|
| 1 | Consultant | 19 | 1 | Yes |
| 2 | Trainee | 11 | 1 | Yes |
| 3 | Consultant | 13 | 1 | No |
| 4 | Consultant | 16 | 1 | Yes |
| 5 | Consultant | 17 | 1 | No |
| 6 | Trainee | 8 | 0 | No |
| 7 | Trainee | 19 | 1 | No |
| 8 | Trainee | 18 | 1 | No |
| 9 | Consultant | 9 | 1 | Yes |
| 10 | Trainee | 11 | 1 | No |

Table 6.1: VUI Participant Information

### 6.2.4 Room Set Up

The study operated out of National Imaging Academy Campus, on a site that regularly admitted patients for medical evaluation and reporting. This meant that we were able to accurately recreate and deploy a reporting room. Participants were presented with a room that was a facsimile of their normal working environment: the room had a desktop computer, 2 computer monitors, a keyboard, mouse and dictaphone, as would be found in any standard reporting office [241]. The desktop computer was equipped to run the same Picture Archive Communication System (PACS) and dictation software that would be found in the participant's standard reporting environment. We then set our research equipment up around these devices as so to mould ourselves *into* the space as opposed to influencing it. We set up 2 cameras - a webcam on top of a computer monitor facing the participants directly and connected to a locally recorded Zoom call, and another DSLR

camera placed over-the-shoulder behind the participant at a 45 degree angle to see both reporter and workstation. These two angles together allowed us to see both the focus of the reporter and how they utilised the dictaphone face-on, and how they interacted with their workspace and screen content. In addition to these camera angles, we utilised a screen recording program that allowed us to examine how a report was made word-by-word in real time for deeper analysis. This screen recording program provided an additional benefit of allowing us to take a frame of the finished report for analysis before the participant moved on to the next report - due to ethical and technical limitations we were not able to export the reports once they were finished, but we were able to pause the screen recording at the moment at which a report was finished to view it in its completed state.

Figure 6.1: Visual Representation of Laboratory Set Up

### 6.2.5 Tools

Participants used an edited version of NUANCE, a Speech To Text software package that is specifically trained to learn and adapt to radiological lexicon, paired with an anonymised set of medical studies on a PACS built by Fujifilm, designed to store and organise health data and records. Both of these software packages are adapted versions of what participants would use when reporting, meaning they were familiar and did not

need training on the systems prior to taking part. The modifications made to the systems for the purpose of the study meant that reports would not be saved or exported locally or to the cloud, and allowed participants to access a corpus of medical studies that were not linked to their account. This allowed us to maintain complete anonymity of patient data and participant reports whilst maintaining integrity of the operating systems, similar to a Wizard of Oz set up. The dictaphone used was a Phillips Speechmike (fig 6.2), also a tool that is approved for NHS implementation and one often found in radiology departments.


Figure 6.2: Phillips Speechmike

### 6.2.6 Session

When setting up and introducing the study, only the author, the Director of the Academy and the technical manager responsible for setting up the equipment were present in the room. The Director and technical manager were present to discuss the equipment changes (discussed below) but would depart after the participant had confirmed that they understood and were content with the set up, leaving only the author present in the room. The participant would be reminded that their diagnostic ability was not being assessed, and that the author was solely examining the interactive process. Whilst the study was taking place, the author would place themselves against a back wall out of the participant's eyeline and remain quiet, only present to ensure that the equipment was functioning properly and that the participant had no questions.

When participants entered our mock reporting set-up, they would be provided with consent and information documentation, as well as a quick briefing on how our system differed from the one that they were familiar with using (such as being locked down with some reduced functionality). The briefing was provided by a technical manager of the site who would also be on hand to assist with any failures or difficulties that participants experienced.

A session would last for 30 minutes, as literature on radiologists suggests that this comprises of, on average, the longest a reporter would sit and continually report on studies with no breaks or interruptions [81]. Once a session had finished, participants would be invited to be part of a collaborative analysis exercise either in person or remotely to view and analyse their own data. We would collect contact details and organise a 1:1 session with participants that agreed to take part – this collaborative exercise comprised of viewing several fragments of video that contained different elements of interaction, with the participant offering commentary and contextual deconstruction of what each fragment contained.

### 6.2.7 Sample Reports

On the Picture Archive Communications System, we had pre-loaded a collection of medical studies from real patients that had been anonymised, with the patient data redacted and any other identifying features removed. This provided us with genuine and appropriate data for reporting, and allowed for us to set up the PACS and Speech-To-Text platforms

properly. The majority of the sample medical studies were Chest and Abdomen X-Rays, comprising of a single still X-Ray image of a patient's lungs and ribs, with 3 CT scans also available for reporting on. The reason for fewer CT scans than X-Rays was practical – when discussing the set up with the clinically-based authors, we came to the conclusion that it would be best to see as many individual cases of interaction in a session as possible, and as CT scans can take upwards of 5 minutes in some more complex cases, we decided that it would be best to encourage participants to review both a series X-Rays and at least one CT scan to provide as much data as possible.

## 6.3 Analysis

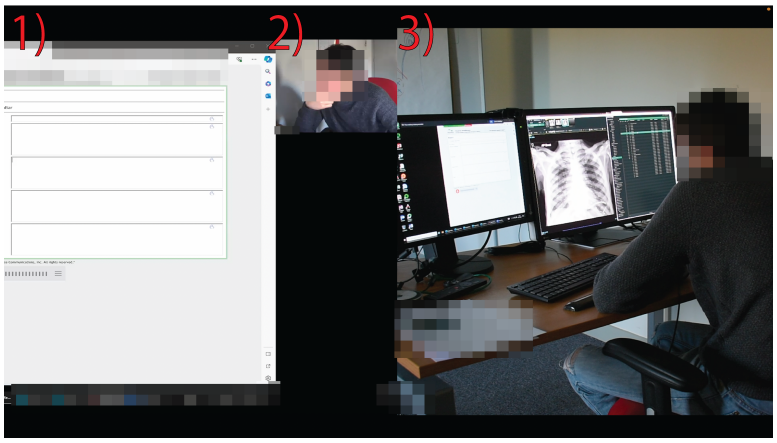


Figure 6.3: Example Of Video Analysis View

### 6.3.1 Independent Video Analysis

This chapter utilised analysis from an independent researcher not part of the general research team, Mx Delahaye. Mx Delahaye was a member of the Swansea University Computer Science department, but did not have experience with health-based research or ethnography - this decision was made to provide an impartial perspective, as the research conducted in this chapter was based upon findings in the previous chapter; Mx Delahaye's lack of relationship to the work was seen as a positive aspect, as it allowed them to make observations and analysis from a completely impartial viewpoint. Mx Delahaye analysed the work independently from the author having only been given the data, and was not involved in the conduct of the research.

Data analysis took the form of 3 key passes: initial viewing, analytic viewing, and collaborative viewing. Video data would be compiled in editing software to show both camera angles simultaneously, and an initial viewing of each session would be conducted by two members of the research team independently to orient themselves with the nature of the data. Rough, unstructured notes would be taken on interaction methods, areas of friction and errors where both human and technology were responsible. Discussion would then take place on areas to focus on, and an analytic viewing would take place that catalogued errors and mitigation techniques, time taken to proof and correct reports, and detailed notes to reconstruct the interaction process. Further discussion would then take place on the most representative fragments to review with participants that wished to take part in the collaborative analysis. Errors were categorised based upon Myer's framework, adapted to suit our specific clinical context, into *"User Input"* and *"Software Recognition"*, and within each category subdivisions of *"Semantic"* and *"Non-Semantic"* [213, 215]. These categories were chosen to cover all aspects of error that have been identified in previous work, but enabled us to code these errors from video footage without prior input from the participant or the system on why or how an error took place.

- User Input refers to the practitioner making a mistake, such as incorrectly commenting on an aspect of a study, or misspeaking when reporting. This included if a practitioner would begin a sentence of analysis and abandon it in favour of a different topic, or realised they had incorrectly identified something later in the report and went back to change it. Software Recognition was categorised as when the participant would correctly vocalise something, but it would be recorded and transcribed incorrectly in the window in front of them.

- Semantic determinism differentiates between errors that would not affect how the report was interpreted or the treatment outcome due to their lack of relation to the content i.e. substituting an entirely different word or phrase and resulting in an incomprehensible sentence: *"8 electricians" to "Atelectasis"* in a report would not be able to be correctly understood by the referring clinician. Semantic error, however, refers to when the system or user would use a word that is incorrect but still allows for an interpretation of the report, and changes the outcome: *"No sign of metastasis"* vs *"A sign of metastasis"*.

- How these errors were corrected would be recorded and sorted into 2 categories of *Vocal Correction* or *Typed Correction*, referring to the modes of text input at the practitioner's disposal - they are able to highlight and re-enter via the microphone, or can choose to utilise their keyboard to type the correction. Reports with noted errors that went unnoticed and unfixed would be recorded, as well as reports that were not deemed by the participant at the time to have any errors in them and as a result were not edited in any way.

Also recorded was the time spent reporting, classed from when the first word was spoken to when the practitioner was no longer analysing the study, and time spent proofing, classed from the end of the reporting time to when the report was erased.

### 6.3.2 Collaborative Analysis

During the collaborative analysis sessions, fragment clips ranging from 15 seconds to 2 minutes comprising of an *event* would be shown to participants over teleconferencing. The author and Mx Delahaye were in attendance of these sessions, and each would have an opportunity to ask questions. The fragments shown to participants were agreed upon by both researchers. Examples of these fragments included a participant reporting on a study without having pressed the record button, a participant correcting the term *Atelectasis* from a mis-transcription by the software of *8 Electricians* and a participant editing a CT report in full using only keyboard as opposed to the dictaphone (as was standard practice). The video would be played several times for participants, and unstructured discussion on the nature of the event, how frequently it may occur in standard daily workflow, and in occasions of error or failure what they believe the cause and solution to be. Unstructured discussion would then take place, whereby the clip would be played on a monitor, and open ended questions would be asked by one of the researchers to the participant in question as so to remain objective, such as "*Can you explain to us what happened there?", "Is that a common occurrence?" "How do you normally deal with a mistake like that?"*. The participant would also be encouraged to offer input on how practice should be influenced based on the data and results they were discussing, including speculative design and pedagogical guidance. Specific care was taken before, during and after the collaborative analysis to ensure participants were aware the purpose was not to expose possible bad practice or to demonstrate the ways in which the participant made mistakes, but instead to investigate the ways in which errors can be made and mitigated for future

reference. Due to scheduling restrictions from participants, these sessions all took place over video conferencing, and were recorded. The resulting transcription will be the source for commentary and quotations throughout our results and discussion sections.

## 6.4 Results

### 6.4.1 Dictaphone Interaction Reconstruction

As in Chapter 5, we will begin the results section with a thick descriptive reconstruction of the interactive process. We will not focus on individual idiosyncrasies in significant detail here, instead offering a collective vision of what constituted the reporting behaviours consistent across the whole participant pool. As seen in Chapter 5 and further reflected here, participants would all have different preferences and standards for the length, verbosity and structure of their reports – as such, we have chosen not to examine these aspects, as we feel it would simply create an extra level of noise and complexity that would not offer actionable recommendations beyond standardised reporting (something that has in itself been the topic of several examinations).

The room and desk is set up as it would be in a standard and generic NHS radiology office: there are a pair of computer monitors both oriented in landscape, with the monitor on the left hand side containing the facsimile NUANCE software and the right hand containing the medical study and PACS software. The NUANCE window is where the report is transcribed, with a series of text boxes each labelled to create a quasi-structured report for the referring clinician. The boxed are headed Title, Clinical Findings, Impression and Plan, but in this situation due to the anonymised and pre-cleaned participants only used the Clinical Findings box, as the lack of previous medical information or a referral notice meant that they did not know what treatment they should be advising for. At the bottom of the monitor is a small digital gauge that crosses the screen horizontally, with a small microphone icon in grey. When the microphone is triggered on for recording, the gauge turns green and begins to monitor the input volume of the radiologist's speech. Input that is considered too low or too high to create an accurate transcription turns the gauge yellow, and input that will cause clipping or damage turns the gauge red.

In terms of the praxis of reporting, the participant pool as a whole followed the same

trends and patterns of behaviour. Participants would settle themselves in front of the middle of both screens so that they were equidistant from their face, and then shift bodyweight and position in their chairs depending on which screen was being examined. They would take the dictaphone in one hand, often the left so that they could manipulate the mouse and keyboard with the right, as is the common layout of an office peripheral setup. Occasionally if the participant was using the mouse often to manipulate the scan, they would input single or small keystrokes with their left hand whilst still holding the dictaphone. The dictaphone would when remain held during the reporting process but in a relaxed position away from the face – when the recording was triggered, it would be brought close to the mouth and participants would speak directly into the microphone at the top of the device, entering the "reporting mode" mentioned in Chapter 5 with low pitched consistent speech patterns. When edits were required that necessitated more than a single key stroke, such as inputting a word or reformatting as opposed to simply deleting an excess space, the dictaphone would be placed back on the table so that both hands could be used on the keyboard. Once satisfied, the participant would resume the original position of dictaphone in left and mouse in right hand.

The actual reporting technique would comprise of focusing in on the medical study and providing a vocalised medical opinion. We went into significant detail in Chapter 5 about the wider reporting process, but the point to identify and reinforce here is the "self-gating" – the recording would only be triggered on for active speech and then triggered off again. Due to the continuous on-off switching of the dictaphone, participants believed there would be less chance of ambient noise creating unwanted transcriptions, but this had the unintended effect of meaning it was on occasion difficult for the participant to remember if their reporting was triggered on or off when their attention was on the medical study. We also saw this in Chapter 5, indicating it is a common occurrence somewhat taken for granted when reporting.

The reporting window is web-based, hosted on a local server that allows for the closed loop sharing of medical information across the health board. As the participant is speaking, there is a brief delay before the words appear on screen, with the longer sentences (i.e. 10 words or more) taking significantly longer and then appearing in full. This meant that when a participant was particularly focus on the medical study, or was giving basic

information that did not require concentration but was memorized such as a description of normal findings, they would have to pause and switch their focus to the reporting window to ensure that the transcription was accurate. All punctuation and formatting is initially done verbally, as both hands are preoccupied with other peripherals. This means that terms such as full stop, colon, comma are spoken as part of the sentence, and new line or new paragraph used to aid legibility and layout. This means that reporting speech differs distinctly from naturalised language, with an example paragraph being

*"Chest new line AP erect full stop the patient is mildly rotated full stop mild proximal thoracic scoliosis comma convex to the right full stop"*

This can, of course, lead to common homophonic errors such as the system writing colon in reference to the body part instead of the punctuation, as well as simple recognition problems of simply inputting full stop as plain text instead of punctuation. It is also worth noting that this only happens in a forward motion, by which we mean participants will move horizontally left to right and vertically up to down when adding punctuation and layout, but will not use verbal layout tools to navigate and correct the whole report. In simpler terms, new line and new paragraph will be used, but not erase line or erase paragraph.

Participants opined that talking with a yellow or red colour on the gauge would lead to more instances of a mistranscription (as would be the intuitive implication) but there was a lack of certainty that this is what the gauge was actually designed for, or if this was simply a folk theory. However, in combination with the significant delays in sentence reconstruction that , this does go some way to address the communicative and interactive techniques identified in Chapter 5: if participants are constantly and continuously being shown a gauge that warns them when speech gets too loud or too quiet, it would implicitly be enforcing a certain volume of input that remains consistent – ergo, we see the emergence of a monotone, lower than speaking volume but with an unusual cadence and structure method of interacting with the VUI that we observed in-situ.

### 6.4.2 Semantic Errors

For the purposes of this chapter, we are defining user error as Reason does in their work on human errors as *the failure of a planned sequence of mental or physical activities to achieve its intended outcome when these failures cannot be attributed to chance* [281], and system error as the malfunction or failure of technology to the extent that the result is an unwanted outcome. Common examples of method errors seen in reporting sessions included not properly triggering the recording on or off, lengthy pauses when speaking (mid-sentence and mid-word) and noises made when thinking resulting in mistranscription, and knowledge errors such as incorrect naming of positions such as left and right or misidentification of anatomy. System errors mostly comprised of mistranscription, where the VUI incorrectly identified words and punctuation – NUANCE software is trained on a radiological lexicon and so had in its dictionary a comprehensive understanding of words that would not be found in common conversation, but would still often mistake common words.

However, the most notable classification between errors was the semantic nature of the mistake made. What is meant by this is the ability to have a mistake that meant that the end report was still legible and logical, but changes the nature and meaning from what the reporting clinician intended. These errors differ from ones that, whilst they restrict the overall legibility of the study, are clearly not the intended meaning of the reporter – for example, a frequent transcription mistake identified in the study was “8 Electricians” transcribed in place of the medical term “atelectasis”. There is no feasible way that a referring clinician could read the end diagnostic report and believe that the participant intended to say 8 electricians when examining a chest plain film. This error was also easily identifiable, as when proofreading the report the reader would be able to notice a mistake like that, as the sentence would be rendered nonsensical. As such, these errors add to the temporal and emotional friction that compounds over a report’s lifecycle of errors, but do not pose an immediate and sentinel danger.

However, we also observed errors that would allow for the report to be read legibly but would result in different diagnostic conclusions from the referring clinician. Terms that sounded similar homophonically but with the opposite meaning held the most concern. Instances of this included mildly recorded in place of markedly, liver recorded for lung, and hypoinflated for hyperinflated. Whilst we did not observe it, participants also offered

in discussion the danger and frequency of no and known – P4 shared an anecdote of this mistranscription error happening to them personally that became committed to patient notes and had an impact several years later, as it was not caught. These errors were well known by participants, and perceived as a key danger of using VUIs as the sole method of interacting with the free text report

*"It's time consuming, and also this patient may have completely different management based on that single word missing from the report"* – P1

When addressing mitigation and reduction, participants had an apparent lack of information for why these mistakes happen, where the fault lies, and how to prevent them from occurring frequently. This meant that mitigation of errors was left to personal beliefs and preferences of interaction, and a trust that they would catch such mistakes in the proof reading process. Participants did opine that they often felt it was their own fault for using words that could be homophonically confused, and claimed that the correct technique should entail avoiding these words

*"It's possibly something I should train myself not to do"* – P4

*"Some consultants will say don't ever use that word, because it comes up wrong"* – P2

*"I have a feeling if I ever managed to get in the habit of pronouncing it differently, it would be less problematic"* – P9

This demonstrates that participants get little assistance, feedback and guidance on mitigating serious semantic errors of transcription that could have significant impact on patient care when interacting with VUIs, and as such the recommendation is simply to change their own language to fit around the precarious system as opposed to vice versa.

#### 6.4.2.1 Proofing Techniques

Due to the total autonomy that speech to text software provides to the radiologist, they become responsible for the total editing of the report before it is sent to the referring clinician. This means that, prior to moving on to the next medical study, they are

required to proofread their work to identify and correct mistakes. We offer some brief statistical information as contextual and supporting evidence, but wish to impress upon the reader that this is for illustrative purposes rather than for in-depth statistical analysis. We will then go into further detail regarding interactional techniques and their efficacy regarding the proofing process.

In terms of time management, proofing a report comprised of 26.7% of the time required to report a CT scan and 28.9% of the time required to report a plain film. Considering that it took on average almost 3 and a half minutes to report on a CT scan (compared to under a minute for a plain film), this indicates that a considerable amount of the reporter's time is spent on the proofing process. P4 claimed that the "complexity of the report affects this proofing time", referring to the fact that CT scans required considerably longer reports due to the level of detail in each medical study.

This extended attention placed on proofing also implicitly points towards the ubiquity and consequences of mistakes in the report. CT reports had an average of 3.67 errors before correction took place, with plain film having 1.46 – again highlighting the impact that length and time may have on the number of mistakes made by both user and system. P1 explained *"it's something you're acutely aware of every time you report"* and P2 stated that they know "I'm going to have to delete things anyway" when asked about the time spent examining the report after its completion. Further, P1 made an offhand estimate that they receive between 6 and 8 emails a year from referring clinicians that note a mistake in a report they have received, but it was evident that this was not of major concern to participants. Errors and mistakes were perceived as unavoidable and simply a part of the process, with P9 stating they would *"rather speak concentrating on the scan, in the knowledge that what's coming up on the screen isn't exactly what I want"*.

In terms of proofing techniques, we identified two distinct methods that participants employed: frequent glancing throughout the reporting writing, and a thorough end read through. Most participants would prefer one method, but would employ both at various stages and depending on the complexity of the report.

Frequent glancing would be made whilst reporting was ongoing. This was usually when long sentences were in the process of being transcribed, where participants would

continuously break focus with the medical study to examine their report for mistakes or errors and correct them immediately before continuing with their analysis. As previously established, participants would be sat with their body facing towards the screen with the medical study on it whilst reporting, but their eyes would frequently glance over towards the reporting window when they had paused for breath or to take a closer look at part of the medical study.

If a participant had experienced an instance of mistranscription or other type of failure during the reporting session, especially one that necessitated a full stop in reporting, these frequent glances would increase in frequency and duration, almost indicating that the participant had lost "trust" in the system and that it required closer supervision for the rest of the session. It is difficult in this study to quantify the overall impact that frequently breaking concentration and removing attention from the medical study had on the quality of the report considering we are not assessing the medical capabilities of participants, but it was noticeable that participants who had (or believed they had) more error-strewn reports adopted it as their primary method of choice for proofing and ensuring accuracy, such as P1 – *"I spend a lot more time checking, because I realized that the reports are a lot more inaccurate"*.

The other method of proof reading was to examine the whole report after it has been completed, from start to finish. Even if participant also engaged in frequent glancing, a final proofread was seen in every participant in the study. For this, the participant would completely change their stance and attention in front of the monitors to station themselves in front of their report, and would place the dictaphone down in order to read. Correction would then take place linearly, as and when a mistake was noticed. This method was seen as embracing the ubiquity of errors, with P9 stating they would *"rather speak concentrating on the scan in the knowledge that what's coming up on the screen isn't exactly what I want"* in contrast to the glancing method.

Most participants preferred utilising one method over another, with only P2 having as many errors corrected during their reporting as they did in the proofing stage after a report was written. Whilst individual preferences could be seen in the disparity of corrections made during and after, there was no discernible preference across the whole

participant pool: across a whole session, we saw a cohort average of 10.1 corrections during and 11.6 corrections after.

### 6.4.3 Correction Techniques

Much like with proofreading, there were two primary methods we observed of participants correcting the errors they came across: re-vocalising and manually correcting with other peripheral devices.

Re-vocalisation comprised of utilising the dictaphone to re-enter an intended word or phrase, or erase an unintended word, by using the mouse to highlight the error and audibly speaking the correct phrase again, replacing it in place without disrupting either the participant's flow or the structure and layout of the report. This method allowed the participant to remain in place and resume reporting, and was supposedly the "correct" way of interacting the VUI system according to P4, but appeared to be error-prone in of itself: we frequently observed that if the replacement re-vocalisation was not recognised correctly again, participants could end up repeating their correction several times, having to highlight and replace each time, resulting in a significant time loss. The re-vocalisation would also on occasion place the mouse cursor in an unintended part of the text box, meaning a participant could speak entire sentences not knowing they were being transcribed in an unrelated section, also requiring further editing. Regardless, this was the most common method adopted by almost all participants for minor corrections.

The other method employed was more traditional keyboard-based interactions with the text. Here, participants would correct as they would in other word-processor type interactions, simply identifying the error with the cursor and erasing/retyping with the keys by hand. This was less common, and often seen as the option to employ when participants had already tried and failed with re-vocalisation: P4 opined that they saw that the need to type "is a failure" in terms of both system and human error. The notable exception here is P9, who utilised the keyboard overwhelmingly more than the dictaphone for correcting at a ratio of 2:18, a considerable outlier. However, we also observed that there were no instances of a keyboard entry being incorrect and needing multiple attempts, in contrast to re-vocalising.

As such, we see that despite re-vocalisation being the primary method of correction, turning to the keyboard was seen as the ultimate "safe option" if previous attempts have failed. Addressing their preferences, P9 stated that "if I just type it, I know its going to be right", indicating the control and safety that a keyboard provided was distinctly preferential in instances of repeated failure. This was also seen as the most useful option when correcting punctuation errors such as new line and full stop, where a single key press would produce the necessary results – breaking from dictaphone interaction was not preferable, but it was reliable.

## 6.5 Discussion

### 6.5.1 "Noise" and Variance In Interaction

When reviewing our results, it was clear that the findings of this study reinforce our theories of interaction in Chapter 5 regarding the uniqueness and idiosyncrasy of techniques and reporting methods, even within a sample size that all work under the same organisation or health board. The radio of semantic and non-semantic errors varied considerably between participants, alongside an inconclusive mixture of user and system fault when determining the origin of an error. To illustrate, P3 experienced 28 instances of a non-semantic error caused by a system misrecognition, with the next highest in that category being 9 by P8 and a cohort mean average of 6.4. This clearly demonstrates some aspect of P3's choices in interaction or inherent language qualities that was causing a significant amount of transcription errors on behalf of the VUI that was unrepresentative of the wider participant pool. Another juxtaposing example is P6 and P8, who reported on similar numbers of scans and had comparable error rates overall, but with P6 making 19 user-based errors but experiencing 4 system-based errors compared to P8 making 5 user-based errors but experiencing 18 system-based errors.

These inconsistencies in system and user performance when using the same data and the same facsimile system indicate and reaffirm the personalised nature of reporting and the impact it can have on digital interactions – the system appears to be generic and offer little practical assistance and feedback to allow for this fluid input and variation in interaction technique. P1 explained that techniques can also be born from environmental factors, and that whilst they believe that the system is generic, the theory was that the way that they set

up their own personal reporting office can impact the performance of the VUI, explaining their "*computer is really set up and tuned so the sensitivity of the mic is just right for me*". P1 did acknowledge, however, that they believed themselves to be an outlier in this fashion and that their colleagues would not put this amount of attention into the setup of their personal office.

Beyond literal patterns of interaction, P2 also acknowledged that stylistic preferences can have an impact on the type and frequency of errors experienced – when analysing an error where the system has misplaced their cursor and begun recording a new line in the middle of a different sentence, P2 claimed that often they "won't report in order", going back through a report and adding comments in different places, often resulting in formatting mistakes. This lack of standardisation in radiology reporting is reflected in the noise in the data gathered, and demonstrates the role personalised methods can have on the number and category of errors experienced within a session – even when using the same system on the same sample reports in a controlled environment, we see a vastly different spread of errors experienced and made.

### 6.5.2 Radiologist As An Auteur

Another factor that arises when viewing our results holistically is that the role of the radiologist goes significantly beyond clinical duties of medical study analysis and expands into roles of typist and editor. Even in a laboratory setting unlike Chapter 5 where the participants faced no interruptions, which they themselves opined was unrealistic, there were a significant amount of unfixed errors that remained even after participants claimed to have fully proofread them. The radiologist had to spend a significant portion of their time allotted for medical analysis on editorial work ensuring that the VUI had done their job correctly, a role previously performed by a medical typist. The key difference here is that, despite the speed and autonomy that VUIs provided to the radiologist compared with having to send audio recordings to a medical typist, the participants in our study overwhelmingly agreed that they did not have in depth training or education in the administrative skills required to be an effective editor of their work.

We describe this as a radiologist adopting an auteur role, a term taken from the creative arts meaning a creative who has full artistic control over a project – here, the

radiologist is responsible for the analysis of the scan, the style and structure of the report (including where actionable findings are and the verbosity of them) and to what extent the report is proofread and edited before it is exported to the referring clinician. They have become the sole influencer and author of their work thanks to the autonomy provided by interacting with a VUI. Digitisation and the resulting increased responsibility of the radiologist has been a topic of debate for several decades, with the common conclusion being that it is, at best, a double-edged sword - the efficiency yielded by digital assistance being offset by a larger number of errors [175, 144, 243].

However, despite the relative ubiquity of VUIs in modern British radiology, it still appears as though they have not been provided the affordances to be full proficient in this role. It would be expected for a trained and hired medical typist to be qualified in these skills, but this is not a requirement for radiologists – similarly, the methods and techniques for interacting with these VUIs is left to the assumption that they are simply intuitive enough to merit full implementation across the medical landscape. If we are to continue encouraging high levels of personalisation, creative control and autonomy over reports with little supervision, there is an indication that VUI implementation must be kept generic to adapt for this. In these circumstances, the radiologist must adapt to develop proper, codified techniques of interaction and error mitigation to ensure a low error-rate. Specific outreach in the form of educational seminars and regular "refresher" events have been proven to be effective in entrenching behaviour and best practice in diagnostic situations [107, 255, 258], and should be an avenue that training organisations should approach to accommodate for the perspectives shown in this study. As such, radiology training should adapt to incorporate elements of digital literacy and competency to ensure sustainable and accurate interactions with VUIs.

### 6.5.3 "One-Way" Interaction

Of note in these circumstances is the way in which interacting with the VUIs at the radiologist's disposable is not a collaborative effort. These are not Conversational User Interfaces, and offer little in the way of feedback or assistance to practitioners when they are reporting and proofing their work. Our participants could not query words that had been frequently mis-transcribed or *look behind the curtain*, nor could they discuss preferences or style like they could with a medical typist, with whom they could form a working

relationship and mutual understanding. Studies have demonstrated that clinicians create mental models of digital systems in order to treat them as they would a colleague [64, 62], but here they are presented with a black box model responsible for the transcription and communication of sensitive and specific medical information.

The primary issue with this singular flow of direction is that participants had little insight into how errors come about and the best ways to mitigate them. Modern clinical decision support tools often come with an element of explainability, but here it is simply left untraceable. This led to significant frustration from our participants, especially in the examples of a non-semantic error – their perception of the VUI meant that they believed it should be capable of knowing the semantic intentions of a sentence, and as such nonsensical substitutions and mis-transcriptions were viewed as a serious annoyance.

*"I can't figure out when it decides to do that... I don't know how it works"* – P1

"[transcription issues] *happen all the time... I don't know how to fix that. It's not as though it hasn't heard what I said*" – P2

"[on a non-semantic system error] *Nobody in their right minds would say that!*" – P9

Several participants opined that they would like to see explanations or evidence for why repeated errors occurred, in order to adapt their performance and technique to the system's tendency to fail in these areas, mirroring what is found in literature [64, 85]. We see that this one-way method of interaction causes frustration to users due to a combination of a lack of codified interactive techniques and a lack of explainability built into the system.

### 6.5.4 Visual Feedback for Diagnostic VUIs

Above, we have highlighted the ways in which interaction with VUIs in these settings is suboptimal – here, we offer a way in which we believe systems can incorporate new design elements that will align with reporting preferences and existing clinical workflow without requiring clinician intervention or adaptation in the form of better visual feedback. Auditory feedback is occasionally seen in radiology departments, such as in Chapter 5, where a participant had a beeping sound inform them if the recording was triggered on or

off, but due to the high ambient noise floor and our observation that participants enjoy listening to music when reporting, here we focus on visual feedback as a primary assistive method.

Currently, the NUANCE software seen in this study operates much like a traditional word processor, offering visual feedback when words have been incorrectly spelled by providing a red underline. In this instance, however, it makes little sense for the system to operate like this, as this is intended to ensure spelling is accurate when utilising a keyboard input – it is impossible for a vocalised word to be misspelled, and extremely unlikely for the VUI to spell a word incorrectly if it is in the system's dictionary. As such, this underline is rarely seen and does not contain much useful feedback to the viewer. However, highlighting of information *has* been shown to be effective in clinical contexts when done by practitioners on their own work, with context awareness [290, 168].

Instead, we recommend developing and deploying ways of having non-intrusive visual feedback that assists with the proofing and editing process that act more like a clinician would. A simple intervention would be to adapt the existing system of highlighting that participants would recognise, but to adapt it to the semantic requirements of radiology reporting: an intelligent, context-driven digital assistant trained on free text reports and medical information in this domain would be able to understand the semantic meaning of a sentence and highlight to users when an inappropriate term had been used by highlighting it, taking out an element of guesswork from participants. Further, the ability to highlight actionable findings or any content that deviates from the standard in a free text report as it is being written would allow practitioners to ensure that the correct recommendations are being communicated to the referring clinician - explicit cues that directly refer to actionable findings are effective at improving task resumption [259, 260].

By introducing highlighting that is contextually specific and focusing in on actionable and communicative results, we would ensure that proper feedback is provided to practitioners but in a format that they recognise, and is unobtrusive. Underlining these findings is a function not available on current VUI systems, but would help to support the radiologist in their auteur role.

## 6.6 Summary

In order to address the research question *"How Do Radiologists Interact with Devices In-The-Wild When Constructing Diagnostic Reports?"* we decided to undertake a study that deepens our understanding of interactions with devices when constructing reports - here, we sacrifice answering how this interaction takes place *in the wild* to more accurately place emphasis on the interaction process with a specific device that constitutes a significant part of the *construction of diagnostic reports*.

This chapter consists of an *ethnomethodologically informed ethnography* that examines the relationship between the diagnostic practitioner and their VUI, the primary means of recording and communicating their domain expertise. We offer a reconstruction of the interaction process, as well as a critical examination of the type and frequency of errors that arise during the report construction. By analysing these errors and techniques in collaboration with members of the participant pool, we offer explanations of where design, training and implementation can intersect to improve the standards and quality of the report construction process. As such, when answering *How Do Radiologists Interact with Devices In-The-Wild When Constructing Diagnostic Reports* we are now able to provide a comprehensive and in-depth reconstruction of the reporting process by combining our observation with this laboratory study.

In Chapters 5 and 6, we establish a considerable amount of significant findings from our observations and discussions with participants, and use them to drive potential design ideas. These design ideas should be further developed and assessed by end-users to discuss both their intrinsic potential efficacy and to evaluate our understanding of what good interventions should look like in this area.

By creating potential interventions based on our own findings and presenting them back to end-users, we can gain valuable feedback on how effective this cyclic process of designing via observation of Work As Done and collaborative analysis is. As such, the following chapter will gather groups of users and practitioners in radiology departments to discuss our conclusions and ideas and provide domain expertise based in their personal and professional experiences.

# Chapter 7

# Design Guidance and Reflections from Domain Experts

## 7.1 Introduction

Throughout the exercises in Chapter 4, participants consistently highlighted the importance of well facilitated communication and digital support. This was later further reinforced in the many mediums of communication participants could utilise when conveying an actionable outcome in Chapter 5. Similarly, our investigation of VUI interactions in Chapter 6 revealed the lack of support that reporters receive from their transcription software, increasing the burden on the radiologist to ensure accuracy. We wanted to explore how utilising these highlighted pitfalls in design ideas would be perceived if presented back to participants in the same healthcare space. Additionally, in all previous studies, we have often offered participants the opportunities to give their own perspectives regarding how existing issues could be mitigated or new novel ideas could be implemented. Chapter 4's design fiction activities offered us 12 aspects of functionality that clinicians would like seen implemented in some way along their workflow, and Chapters 5 and 6 utilised semi-structured interviews and discussions to give radiology practitioners an avenue to present their personal workarounds for areas of latent and persistent friction.

Here, we explore some of the commentary and perspectives that have arisen over the previous chapters; we conduct a pair of focus groups with clinicians and technical managerial staff in the NHS investigating their perspectives on potential designs that were based upon

suggestions from participants in previous studies, comparing them to existing systems and tools that they are familiar with. We then engage participants in speculative design fiction after exploring potential designs similar to what was conducted in Chapter 4. We designed three probes to present to participants based on this feedback; **Recall** based upon an existing function within PACS that participants had mentioned but we had not seen used in-situ in any context, **Traffic Lights** based upon Do-Not-Disturb functions on other mediums of communication translated for a radiological environment, and **Semantic Highlighting** based upon specific participant feedback in Chapter 6. All of the prompts were designed in FIGMA, and presented to participants as static graphics on paper for them to gesture with and use as examples throughout the focus groups. By having participants evaluate these designs, we can align their experience and tacit knowledge of how work is done with reflective design based upon their opinions of how to address points of friction and failure. Put simply, we are presenting their ideas back to them under the guise of our own, to examine how they would critique them as outsider suggestions. The goal of this is to both to evaluate how effective utilising participant perspectives in this way is, and to add a further dimension to our speculative activities carried out in Chapter 4.

We found clinical participants to be particular concerned with implementation of systems as opposed to possible functionality, with their primary focus being on un-invasive methods of improving their workflow. Technical staff, however, had a negative view of their user base and emphasised the importance of having systems that required little oversight and interaction from them. Overall, participants did not have overly positive opinions of the prototypes and designs presented to them, and instead desired more collaborative "co-pilot" assistance that would not interfere with their existing workflows.

## 7.2 Method

### 7.2.1 Approach

As with Chapter 4, this study took the form of a 3-part focus group on Imaging Academy Campus, with the approach chosen closely mirroring the guidance of Steward and Shamdasani [284], and Breen [56]. As such, we adopted a conversational tone and emphasised the outsider aspect of conducting research in a safety critical domain. We again recorded both sessions with a digital recorder pointed out to participants then placed out of eyeline

on a table, but in terms of design we diverged from Chapter 4 by preparing our own materials for participants to discuss, as opposed to allowing them to design their own.

We added an additional dimension to the way in which we hosted this focus group, by regularly informing and discussing participants of previous studies; in our script and when conducting the focus group, we would discuss with participants our findings in Chapters 5 and 6, and how they influenced these designs and our perspectives on them. We would introduce anecdotes of interesting factors such as how voices adapt to interacting with VUIs, how mistakes are identified and corrected, and how our views as computational experts coming into a clinical may differ from theirs as domain experts. This established our own levels of knowledge and competency as focus group leaders, allowing participants to more freely discuss specialist terms and scenarios, but also allowed us to discuss previous findings with a new participant pool to add depth and context to them.

### 7.2.2 Participants

Unlike previous studies carried out in this thesis, here we decided to include non-clinical practitioners to examine other perspectives on how these systems were "implemented" and maintained – these factors were often brought up by previous participants when examining holistic workflow or tool design, that often the system itself was not a problem but the way that it has been integrated into the workflow is the reason for their dislike. As such, we wanted to examine perspectives "from the other side".

We conducted a first focus group with eight consultant radiologists, but also conducted a second iteration of the focus group with eight IT managers. These managers were from the same health boards as the consultants, and were technical experts responsible for the procurement, operation and maintenance of digital and hardware systems in the radiology departments in which our consultant participants worked. These participants were pre-sorted into these groups, and as such we did not collect any demographic information about them prior to participating in the study.

| Participant | Occupation | Group Number |
|---|---|---|
| P1 | Radiologist | 1 |
| P2 | Radiologist | 1 |
| P3 | Radiologist | 1 |
| P4 | Radiologist | 1 |
| P5 | Radiologist | 1 |
| P6 | Radiologist | 1 |
| P7 | Radiologist | 1 |
| P8 | Radiologist | 1 |
| P9 | PACS Manager | 2 |
| P10 | PACS Manager | 2 |
| P11 | PACS Manager | 2 |
| P12 | PACS Manager | 2 |
| P13 | PACS Manager | 2 |
| P14 | PACS Manager | 2 |
| P15 | PACS Manager | 2 |
| P16 | PACS Manager | 2 |

Table 7.1: Guidance Focus Group Participant Information

### 7.2.3 Session Structure

*Part #1 – Existing Systems*

As our warm-up¬ session, we explained the concept of "usability" to participants using the International Organisation for Standardisation framework, as we felt this would give us a consistent metric to explore all of the following designs and systems against. The framework consists of 3 key aspects, effectiveness, efficiency and satisfaction as measures to judge how usable and system is overall. Effectiveness refers to how well a design completes the intended task, efficiency refers to ease of use and the time needing to be spent learning, and satisfaction is the user's overall attitude and feelings towards the design.

After briefly explaining these metrics to participants, we offered them a series of designs that they would be familiar with in the workplace and in the home, and asked them to give their perspectives on how usable each design was. The intention here is to get participants thinking about these metrics critically, but also allowing us as researchers to understand how they felt about existing systems before presenting them with novel ones.

The way we invited participants to discuss these existing systems was by providing a rough visual Likert Scale, comprised of a heading (i.e. one of the 3 aspects of usability) and a stepped line underneath that participants could visually "rank" designs against each other. We wrote a collection of devices, software and tools on Post-Its, held them up so the room could see, and then opened the room to conversation about where each design lay along each Likert Scale. The existing systems were identified from Chapters 5 and 6, and participants discussed Microsoft Software Teams and Outlook, NUANCE speech to text software and their dictaphones, Picture Archive and Communications Systems, and WRIS, software used for scheduling scans and data storage.

*Part #2 – Improving Systems*

After participants had spent time discussing existing systems, we presented them with our 3 design prompts and explained each one. These designs were intentionally kept simple, as we believed this would allow participants to discuss potential improvements or additional functionality - we felt that offering a simple design prompt allowed participant to imprint their own ideas and perspectives, whilst remaining on task of providing assessment and feedback. We passed out several paper copies with a design on it, and asked participants to spend some time considering possible use cases and situations in which they think it may have come in useful. We then presented participants again with the 3 metrics of usability and asked them to consider how these systems may be implemented in their own clinical workflow, for better or for worse, and the factors that may be associated.

We then asked participants to reflect on their personal experiences and how these design prompts either aligned or diverged from the existing systems we had previously discussed – here, we wanted participants to engage in discussion with their peers about systems specifically designed to meet their needs as opposed to more generic systems adapted for radiological use. Our **Recall** and **Semantic Highlighting** functions were based upon the existing PACS reporting window seen in Chapter 6, and worked to overcome issues of proof reading and editing highlighted in both our analysis and co-analysis with participant researchers. The ability to have a report read back to the radiologist was something mentioned as possible in Chapter 5 when discussing the practicalities of working with PACS, but it was emphasised (and not clarified) that it was not an efficient or common

method of proof reading – we wanted to explore whether this was due to the apparent group habits and folk theories identified in Chapter 5, or because it was impractical.

The **Recall** design (fig 7.1) presented to participants was simple, simply comprising of a PACS text box with a sample report, large microphone icon, and button inscribed with *Recall*. We explained to participants that the design would work by reading aloud the report in a voice similar to a home digital assistant, like an Alexa or a Google Home, something we believed would be familiar to them [80]. We informed participants that the recall function button could be pressed with a mouse click, and would simply read the whole mass of text out loud, but that they could highlight stretches of text and have them specifically read aloud to them. We explained the motivation behind the design as the ability for participants to have their reports read back to them by an external system, lowering the likelihood that they skip over an error in the skimming process of proof reading. The additional purpose of the design was to prompt participant to discuss the reasons that they didn't use the existing function under the guise of assessing a novel implementation along the same lines.

We decided to begin the activity with a design that was a facsimile of one that exists in the real world as a way of introducing participants to the aim of the exercise slowly, and to encourage good quality and fair feedback - we had seen in Chapter 5 and Chapter 6 that participants were willing to give honest and critical assessments of the tools they used when constructing reports, and so we believed that by providing them with this **Recall** design we could begin the exercise with some feedback based in personal and professional experiences.

**Semantic Highlighting** (fig 7.2) was suggested in Chapter 6 by a participant as a way of ensuring higher levels of accuracy without being invasive to existing workflows – the participant opined in the co-analysis session that they wished to see something akin to the underlining function found in word processors and search engines to highlight potential semantic and transcription errors that may change the way in which the report was read by the referring clinician. Copying 3rd party designs that were familiar and intuitive to participants outside the workplace is a method advocated by Militello as a way of mitigating the lack of digital literacy often found in hospital sites [202]. In addition,

we wanted to cover the commonly identified factor that reporting is personalised, and heavily dependent on the style of the radiologist, meaning that the referring clinician is focused upon actionable outcomes and will often not read the whole report before making a judgement.

Here, we again presented participants with a sample report in a PACS text box window, but in this variation we removed the recall button and instead underlined each *abnormal* in the report with a thick red line, defined by any aspect of the medical study that the referring clinician would deem to be out of the ordinary expected results. This resulted in a report that highlighted to the reader anything that would require addressing, allowing participants to judge how effective a non-invasive intervention would be in a realistic setting like this. By introducing a colour found in work processors to highlight an unknown input (such as a spelling error), we increase the likelihood that these sentences would be found in a brief "skim" of the document by the proof reader [168, 290, 4].

We explained to participants how the design would work - we included a button labelled "Summary" in each text box, and told participants that upon the summary button being pressed, the design would scan the text in the relevant box and provide red underlining on any information that was deemed by the fictional and contextually aware algorithm to be an *abnormal finding*. We emphasised to participants that the design would run on a machine learning infrastructure that enabled the design to know *in context* what an abnormal finding would be, meaning only relevant information would be highlighted to participants. On the actual design presented to participants, this red underline was based on the recommendation of the author of the fictional report, a consultant radiologist, meaning it was in the correct and appropriate places for maximum accuracy of assessment. Additionally, we used a different report from the **Recall** design, with the fictional report having a balanced mixture of normal and abnormal findings as so to demonstrate the design working in a legible way.

**Traffic Lights** (fig 7.3) was again based upon conversations and observations of practitioners in-situ. We discuss the high volume and disruptive nature of being interrupted in Chapter 5, addressing the fact that practitioners have to address telephone calls and knocks at the door immediately as there is no way of telling how important the disruption is

before answering it, the implication of this being that relatively unimportant interruptions cannot be avoided even during complex cognitive tasks. In terms of addressing these interruptions, William's MSc project on cognitive loading in diagnostic situations found that priming an individual to the fact that there would be an interruption was effective in reducing error rate and ease of resumption of the task [315], and work from sociologists such as Bell, Watura and Banzinger found that vetting or triaging in these events, adding an additional stage that categorises the importance of an interruption, has significant positive impact on the negative effects of interrupting a complex cognitive task [34, 23, 311]. Finally, Borojeni showed that specifically lighting cues worked effectively at improving task resumption after an interruption [48].

As such, we offered a design that covered both aspects of priming a practitioner to an incoming physical interruption (such as a knock on the door), whilst empowering themselves to vet the interrupter before accepting; traffic lighting was a simple GUI design intended to be hosted on the radiologist's computer monitor. The design consisted of a red, yellow and green light with a button that allowed for accepting or declining. The explanation given to participants was that a similar series of buttons would be outside their office, and potential interrupters would be told to broadcast the importance of their request – emergencies would be red, whilst social or casual would be green. When triggered, the reporter would receive the corresponding light alert on their GUI, and would be able to accept or reject the interruption, signalling the interrupter to come in or come back another time. We reasoned that practitioners divide their time between constructing reports, a cognitively taxing and complex task that requires total severance and resumption, and menial tasks such as organising their desks and replying to messages, that could feasibly be interrupted for even minor reasons.

*Part #3 – Future Systems*

As a final activity, we again engaged in a design fiction activity similar to Chapter 4. We did not provide a fictional scenario here due to time restrictions, but instead simply opened the floor to discussion for how participants would approach developing their own design prompts similar to the ones we had just presented to them. We encouraged participants to focus on a shortcoming or point of failure they had previously discussed in

the first two parts of the session, and offer their own perspectives on what they believed could be implemented or developed to address it. The purpose of this activity was to see how the perspectives of clinicians differed to those of technical managers, to investigate both digital literacy impacts and how being on the other side of maintenance and operation of tools and devices had an impact on a light design fiction.

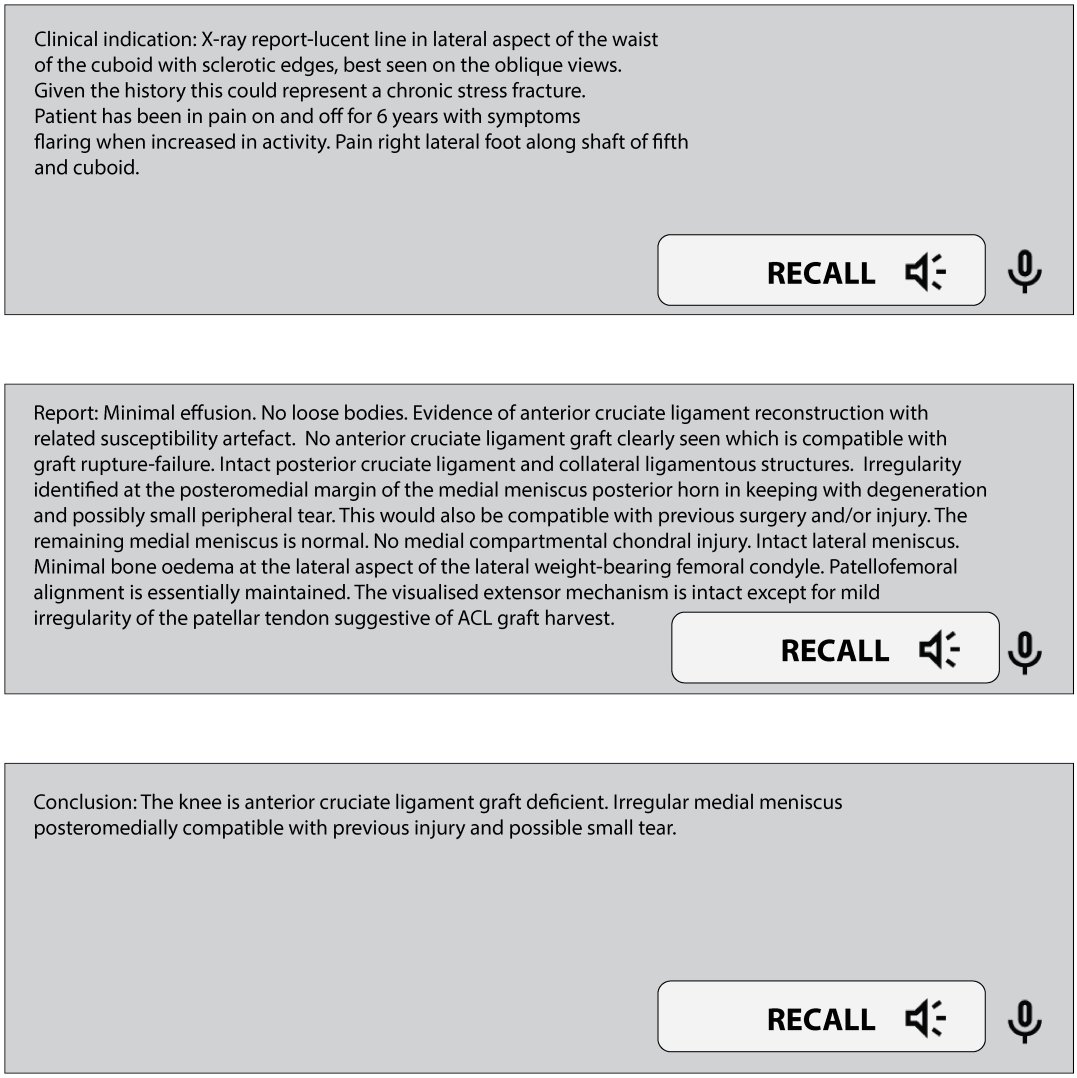


Figure 7.1: Design Mockup 1, Showing The *Recall* Function

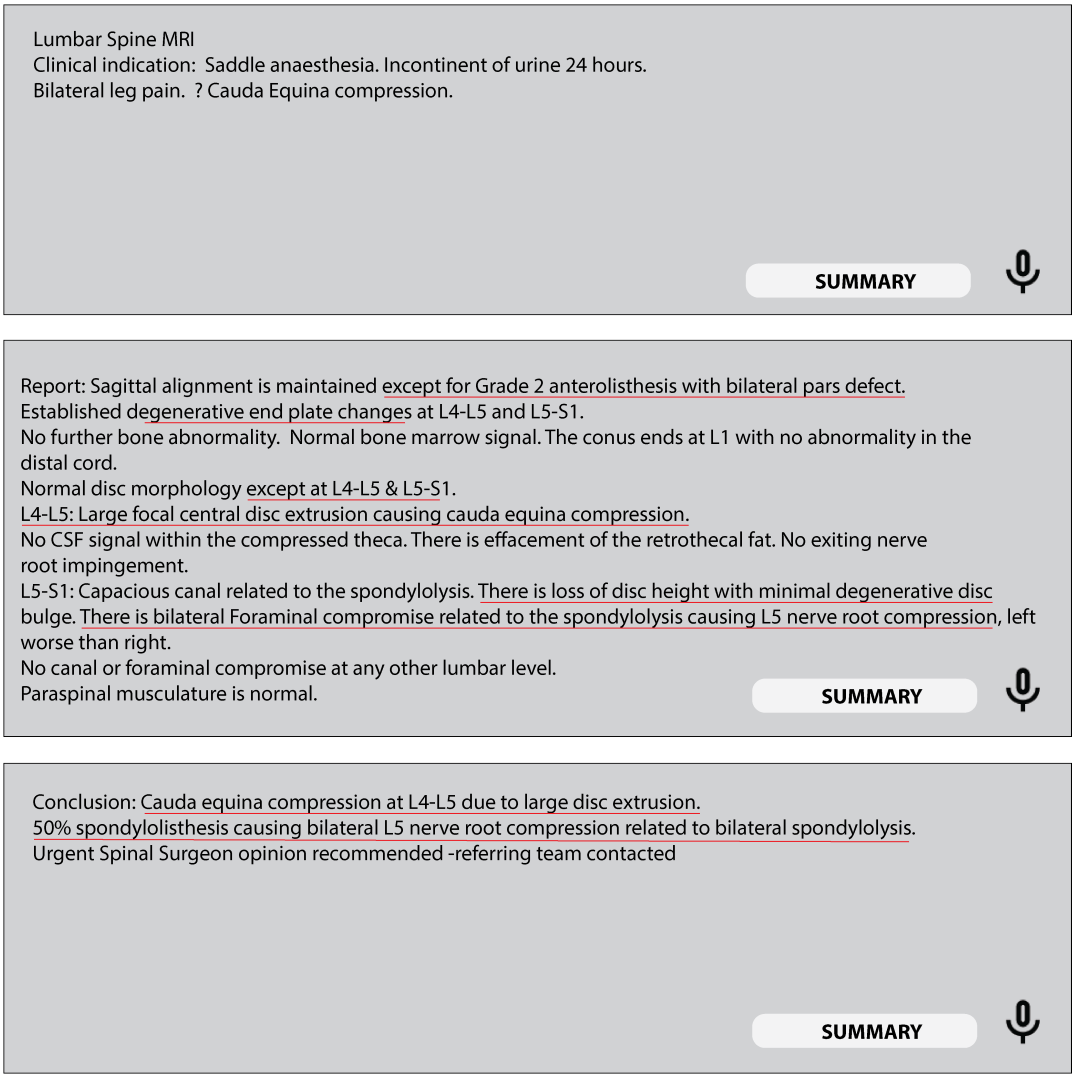


Figure 7.2: Design Mockup 2, Showing The *Semantic Highlighting* Function

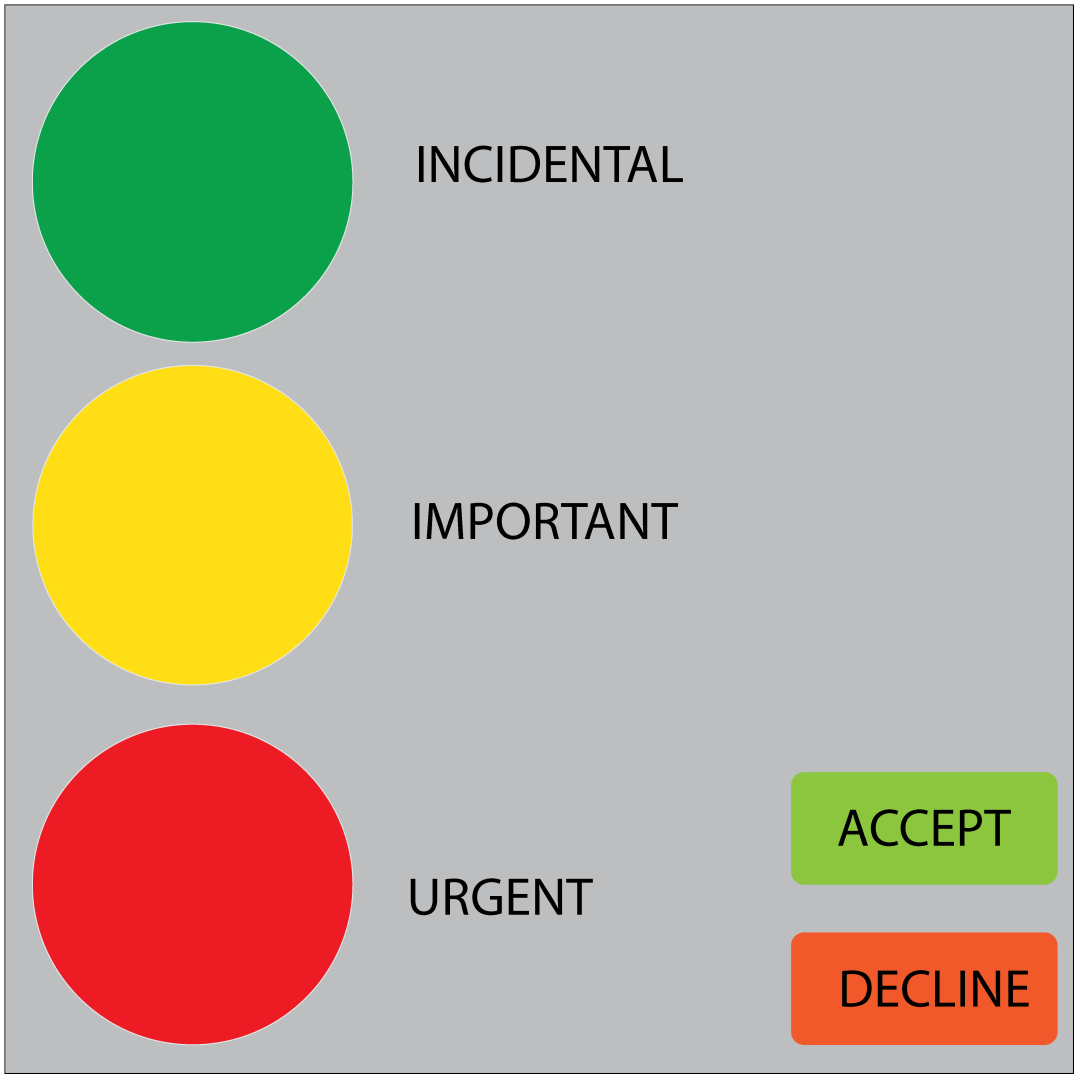


Figure 7.3: Design Mockup 3, Showing The *Traffic Lights* Function

### 7.2.4 Conducting the Focus Group

In both instances, the focus groups were conducted solely by the author. Each group would be sat around a large table in a conference room on NIAW campus while the author would conduct the focus group from the front of the room. Both groups were aware of the author's role and research, as by this point in time this was the 3rd study that had been conducted on NIAW campus. Additionally, having spent time observing radiologists as part of Chapter 5 and 6, the author was informed about practices and key terminology, meaning conversation about topics could incorporate a rich level of detail

and tacit knowledge - as such, participants were familiar in tone and open to discussion about all facets of their role.

In each focus group, participants would be informed that they were one half of the study cohort and that the other was managers or clinically facing (dependent on the group), but that commentary would not be passed between the groups and would be kept anonymous. As a result of this, a considerable amount of discussion was focused around that "the other" group would think about the prompts and discussion points that were taking place. Efforts were made by the author when conducting the study to ensure that the topic was centred around what each group themselves thought, but the results will reflect how the tone was often shifted to systemic and political conversation.

Finally, as was the case in Chapter 4, each focus group consisted of a cohort of participants that worked within the same geographical or organisational remit - this meant that participants were familiar with each other and were comfortable and confident discussing between themselves. As a result, little effort was required on behalf of the author to stimulate discussion or encourage individuals, as contribution was even across all participants.

## 7.3 Results

Both focus groups were recorded digitally, with the audio being transcribed by hand and subjected to thematic analysis by the author. Thematic analysis inspired by Braun and Clarke was conducted, and comprised of three read-throughs each with a reflective aspect to meet the 6-point framework; one for familiarisation with the content, one to establish codes and extract key quotes from participants, and a final time to codify full themes across the whole session. Each session was transcribed and analysed independent of the other, and as such the results will discuss them separately before addressing overarching themes and implications in the discussion section [55].

### 7.3.1 Clinical Group

*Existing Systems*

In general, clinicians spoke favourably of existing systems in terms of fundamental design, but were instead critical regarding the implementation of peripheral hardware

and software. When given the example of Microsoft Teams, participants opined that the software worked as intended and designed, but the poor quality of the internet connection and the microphone and camera provided to use on Teams were the reasons for friction.

*"The software itself is good, the internet and microphones that go with it [means that] every meeting is a disaster"* - P6

*"The environment is poor because you're hot desking... there's not enough attention put into the infrastructure"* - P2

The same was discussed when the example was NUANCE, with participants highlighting personal experiences where the setup and settings of the software meant that it would default to unwanted presets that would be inappropriate for the setting in which the software was deployed, such as the microphone sensitivity being very high in a busy crucible of reporters, letting extraneous noise cloud the microphone input.

*"I don't think the environment has been thought of within the set-up of the workstation"* - P3

These issues in implementation were seen as important because they would have knock-on effects for the ways in which interaction was shaped by newly qualified radiologists; comments were made that more junior trainee and registrar radiologists would be more open to typing reports with a keyboard than speaking them if these problems with implementation continued, something seen as less efficient or effective than their preferred methods.

*Design Prompts*

When discussing our prompts, the common thread among participants once again was that they would not suit the practicalities of working as a radiologist in a standard NHS environment.

The Recall function was widely seen as too time-consuming – after some discussion

it was recognised as something that practitioners had to option of using in the real world, but participants felt that having to listen back to several paragraphs of writing would slow down their process to the extent that it would become frustrating, and that the payoff of catching some potential mistakes did not outweigh this burden. Additionally, participants highlighted the need for higher quality peripheral devices when utilising an audio-based function like this; most of the participants claimed that they have to spend time working in communal spaces, without headphones or proper speakers, and so attempting to listen to the report being read back to them would be ineffectual. Whilst the overall feedback to the prompt was negative, it did answer our earlier questions as to the lack of use of this function in-situ.

Participants were also largely critical of the Traffic Light design. The prevailing opinion was that it would be victim to abuse from junior members of staff and trainees who, in the words of the participants, would believe that everything deserved to be addressed under the red light. This would render the system useless as it provided other members of staff an opportunity to highlight any potential issue as of severe importance. This perception was not wholly conjecture – personal experiences from participants with the red flag system for highlighting medical studies that require a second opinion or immediate addressing by a consultant showed similar problems. The claim was that referring clinicians would simply signal everything with the flag so that they received an answer quickly. Here, participants instead believed being interrupted was a part of the job, and a digital intervention would not help.

However, the Semantic Highlighting prompt received positive feedback. Participants opined that this was something that could more realistically be implemented into their existing workflow, and were more excited that they were being shown a prompt that did not mirror their previous experiences with other kinds of tools. The idea of a non-invasive tool that they had the option of turning on or off at will was seen as the peak of what could improve their efficiency. In their discussion, participants offered design improvements and additional functionality, mostly revolving around the importance of contextual understanding from the fictional algorithm; they wanted to make sure that the design would be adaptive to their personalised reporting styles, and would only extract actionable findings – the critique of the design prompt offered was that, in our fictional sample,

there were too many out of the ordinary findings and as a result much of the report was underlined in red. Participants felt that the underlining would need to be seen sparingly, as so to flag up more serious findings as opposed to a more casual spellcheck and proofing tool.

*Future Systems*

Finally, when engaging in design fiction, participants chose to use the semantic highlighting prompt as a "jumping off point" to discuss functionality of systems they liked in casual use settings and how they could be transferred to a radiology reporting workflow. Many of the keywords used by participants centred around organisational factors such as management and politics – terms such as "privatisation" were used as a comparison point between the NHS way of operating and a more profit driven model, seen as more efficient and enjoyable to our group of participants; other words frequently used such as "accountability" and "project management" again referred to the perception that implementation of systems and tools was the key hindering factor in their usability as opposed to inherent issues with design, a consistent thread through the whole focus group.

When discussing features that participants found useful in non-clinical settings, aspects such as 24/7 technical support as would be found with larger corporations was seen as a key goal, as hospitals do not function on a 9 to 5 timeframe and the lack of consistent support was seen as a frustration with the delicate and complex technical systems in use such as the reporting workflow. Additionally, an "all-in-one" design philosophy was readily adopted by all participants over the course of the final session – P2 claimed that since

*"a lot of us do hot-desking, it should follow us around the building"* - P2

referring to features such as cloud-based support, app-based integration, and access to their data and files anywhere on-site. This was in part based on the belief that local peripherals were lacklustre, and participants wanted the ability to move between devices and departments to find the most appropriate and high-powered devices to use and still have their personal profiles available to them. This was the note that the session ended on, emphasising the key issue of poor implementation and peripheral devices, and highlighting the desire for ubiquitous access to their workflow.

### 7.3.2 Management Group

*Existing Systems*

When discussing existing systems from a management, maintenance and implementation perspective, our management participant pool has 3 views across the group and the session; "Does the job"; "Increased Functionality = Decreased Effectiveness" and "Holistic Evaluation of Systems".

Does The Job refers to a relatively casual and non-committal response from participants when assessing the given existing system prompts. Participants would start by describing the functions that the system was designed to achieve, such as "sending messages" or "viewing medical studies" and consistently agree that, at their core, these systems fulfilled requirements. The response "it does the job" was the most common phrase used when discussing the usability Likerts, and following these there was little discussion in-depth about how well these requirements were fulfilled, simply putting all of the prompts in the "good" column of each scale - there was an apparent goal to be "fair" to systems, and not want to simply critique them for the sake of it.

Increased Functionality = Decreased Effectiveness was a general discussion point after the whole Likert activity had finished – once participants had seen all of the Post-Its clustered together in the same part of the scale, the conversation moved towards comparing various iterations of each tool instead of the tool as a whole. Participants would discuss older implementations of PACS or Teams and assess the various changes that each new update would bring, with the overall perspective of the group being that each new iteration would add additional features that were designed to be competitive with other systems, and as such would detract from the original function, thus lowering the overall usability. For example, when discussing PACS, features in the newest versions such as 3D rendering, AI co-piloting of analysis and an instant messaging function were unnecessary for a system designed to view medical studies. Participants opined that having individual systems designed for a particular function was more likely to lead to a higher quality end-goal being reached.

Holistic Evaluation of Systems was the final point discussed from the activity, after

each tool had been assessed individually. Participants opined that the evaluation of existing systems individually to a scale was an inappropriate way of assessing overall effectiveness, since clinical workflow worked to pull all of these systems together in tandem to achieve an end goal. This discussion point again referred to implementation, but here participants wanted to discuss extraneous factors such as staffing or data storage facilitation – the conversation was centred around how the system would work in-situ as opposed to how effective it was at achieving a singular goal. It is worth noting that this would be covered under the efficiency aspect of the tool based upon our ISO briefing given before the task was undertaken, but it was felt to be preferable for participants to continue discussing this factor amongst themselves than interrupting to clarify the ISO framework.

*Design Prompts*

Overall, the perspective provided on our design prompts by the management participants could be summarised with P12's statement:

*"I think it's good, but they just won't use it"* - P12

There were a variety of reasons behind this statement, with participants citing poor digital literacy among clinicians, complaints of additional functionality and overly complex systems "getting in the way" of a good workflow, and the perceived importance of speed when reporting being above accuracy.

Addressing recall, participants again stated that it was an available function to reporting radiologists but made clear that they felt it was significantly under-utilised for its potential in catching errors. Here, much like with the clinician led group, the reasons behind its poor adoption mostly related to the time required to listen back to a report as opposed to skim reading it. Even short reports such as x-rays that could take less than a minute to complete would have their overall reporting time significantly extended, something management participants opined reporters would be unwilling to sacrifice.

There were also concerns that listening back to an automated voice (the example drawn was with Siri or Alexa) would lead to auditory fatigue, and that reporters would begin to

ignore the dialogue and miss details just as much as reading long paragraphs did. The example was given by P14, who claimed that when they were training to report on medical studies (before they became a technical manager) they would listen to themselves giving example reports via a digital tape recorder for practice – P14 claimed that after listening to a couple of reports in a row, they would begin "zoning out" and would miss if they had made a mistake. The perspective was that, over the course of a 3-hour reporting session, eventually utilising recall would become ineffective.

Similarly, there were concerns over the real-world practicalities of implementing semantic highlighting; participants opined that radiologists had a particular dislike for systems that were invasive, and having considerable amounts of text potentially obscured by red underlining was described as "getting in the way", a cardinal sin of clinical workflow systems. Despite only highlighting actionable outcomes that would require attention, managers described the highlighting as "overkill", and that it would eventually be ignored much like the recall would be. Participants did comment that, if it was used perhaps more sparingly, it would grab more attention – the highlighting would need to be off by default and turned on in order to proof at the point of submission. Whilst this was less of a factor than with recall, it was also outlined that checking every underlined comment would result in a longer reporting time, something they felt radiologists would perceive negatively. Much like with recall, the semantic highlighting design was seen as interesting and potentially useful in an ideal environment – instead, it was the preferences and habits of radiologists and reporters that would hinder their use in-situ.

Finally, traffic lighting was also seen as unsuitable for real-world implementation. The reasoning was aligned with the clinician group, that it would be the victim of abuse outside of its intended purpose. However, the management group approached the issue from another angle – they claimed that, given the option, radiologists would simply reject any incoming notification regardless of the importance signalled by the interrupter. The claim centred around the understood reality that radiologists are constantly interrupted: due to the fact that interruptions were seen as ubiquitous and unavoidable, given the chance all radiologists will decline an interruption that they can to catch up on the work delayed by other interruptions. This cycle contradicts the clinician-led group, who claimed it would be other members of staff who would not use the system fairly. As with the

previous 2 prompts, the group consensus was that it was not the design but the user base that rendered it ineffectual for implementation in the real world.

*Future Systems*

When addressing the design fiction activity, our management participants mostly decided to focus their efforts on ways of mitigating the "they just won't do it" attitudes and preferences of clinical practitioners as opposed to physical design ideas. Participants discussed the need for better digital literacy amongst highly qualified clinicians through regular training and education – the consensus among the group was that it would be easier to adopt and maintain our earlier design prompts if clinicians were more open and perceptive to digital interventions in their workflow.

Alongside this, participants did discuss a function they would like to see implemented across all aspects of their working environment in voice-driven assistants. Using a comparison point of an Amazon Alexa, P13 described a device that would allow for practitioners to access data across systems and platforms in a "middleman" function that allowed to the previously mentioned variety in dedicated systems without the need to be switching across programs and hardware.

## 7.4 Discussion

### 7.4.1 Guidance for Tool Development

Overall, participants did not reflect kindly on the design prompts we provided them with. For the most part, this was not due to glaring issues with the prompts themselves, but simply that they wouldn't "work" when implemented in a realistic environment. From a clinician's perspectives, the demands were simple, comprising of 2 dimensions:

**1) Be Non-Invasive**

**2) Don't Slow Me Down**

These factors came into discussion throughout all 3 phases of our clinically-led focus

groups. The design that was viewed most favourably by clinicians was the semantic highlighting, due to the fact that it required no input from the user to work and could be ignored if preferred – this demonstrates how highly these radiologists value their time when reporting, but also that they have entrenched ways of analysing and communicating reports that they were loathed to changed. As such, from a clinical perspective new tools and designs should follow the rules of being non-invasive and should not add more time onto an already constrained process. However, the management group (despite not using these tools themselves) viewed semantic highlighting as negatively as the other designs. The overall perspective of "this is good, they just won't use it" permeated every prompt provided, demonstrating a divergence between the two groups based on perception.

When it comes to comparing how end-users (in the form of our clinical participant pool) and technicians viewed the same prompts, we see that both have a visibly negative view of the other; the radiologist group claimed that tools were up to standard, but the way in which they were implemented and managed was the reason for points of failure and a lack of satisfaction when using them, and the management group took a dim view of the low digital literacy and lack of enthusiasm for new systems displayed from clinicians.

This 2-dimensional perspective of the same issue, even with a similar outcome i.e. a lack of adoption, provides a more in-depth grounding of the issues we have previously investigated in Chapters 4, 5 and 6; a common claim that arose in these studies was that the reasons for failure in systems was due to poor implementation or low quality systems – by discussing this with technical managers, we can see that inconsistent digital education and resistance to novel developments may also play a role in the same event of friction and failure.

### 7.4.2 Reflections On Methodological Choices

The need for this focus group partially arose from our choice at the outset of the thesis to involve the participant and end-user as much as would be possible, due to the methods utilised in Chapters 5 and 6 with our specific ethnomethodological lens that placed emphasis on *work as done*. By focusing so heavily on the observable and the tangible, it meant that we needed to include participant experience and tacit knowledge to ensure that our results were generalisable.

This approach seems at odds with the feedback we received on the design prompts we presented to members of our research community; the logical assumption would be that by adopting an approach that was heavily focused on pragmatic and real-world work, we would create designs that are directly implementable and beneficial for in-the-wild application. Instead, we were told that the designs would suffer from poor adoption and were unsuitable for use in a real-world environment due to the ways in which radiologists interacted with their social and professional surroundings (such as the *traffic light* system potentially being the victim of abuse by junior and training members of the clinical staff). Our radiologists were, in a way, a fickle group who did not maintain what they wanted from a design from study to study.

This presents a dimension to utilising a *work as done* and heavily participant-led approach that was unexpected at the outset of this research. It becomes apparent that

1. Simply because participants were involved in the analysis of an interaction and the resulting design recommendations does not mean that the end design will be looked upon favourably by the demographic as a whole, nor does it mean that the design will be entirely suitable for in-the-wild adoption

2. Adopting a purely ethnomethodological approach to understanding interaction with an intent towards producing more actionable and effective design guidance is not entirely effective, as the demographic being observed can be inconsistent in what they believe will work as a solution to observable problems

As such, when reflecting these findings back through Chapters 5 and 6, we can see that our intent to reflect participant needs by utilising a "*checking in*" method when conducting ethnographic studies resulted in a greater depth of understanding in the interactions that were taking place and the reasons behind them, but the design recommendations from participants as to how to solve the problems that arose require more investigation.

## 7.5 Summary

This chapter works to address aspects of the question *How Can the Perspectives of Radiologists Be Utilised to Meaningfully Impact the Design and Implementation of Devices In Safety Critical Environments*? Here, we examine the efficacy of drawing heavily from feedback from

participants after conducting research in Chapters 5 and 6 to ensure that our findings remained grounded in contextual and tacit understanding of the *work as done*.

By leveraging a focus group comprised of presenting participants with design prompts based on both our own findings and their perspectives on them, we explore the benefits and drawbacks of heavily relying on participant-led discussion when examining the design of the devices that radiologists use. We demonstrate the ways in which utilising their feedback can result in meaningful potential prototypes (*semantic highlighting*) demonstrating the effectiveness of employing expert users' tacit knowledge when presenting discussions and findings. We also, however, demonstrate that this is not a perfect system and that participant feedback should not be taken as the sole source of data when designing and presenting potential prototypes for implementation.

The focus group marks a "full circle" evaluation of perspectives and devices from Chapters 4, 5 and 6; we began in Chapter 4 by establishing a baseline of clinician-led perspectives on workflow, conducted independent and in-depth analysis and presented commentary in the form of design prompts back to participants to examine the process of forming design rules and frameworks based on this semi-collaborative process. We found some new rules to adhere by, primarily about the ways in which clinicians dislike invasive tools and digital programs, and received feedback on what we believed would be novel implementations to solve perceived issues. As such, when answering the question of how these perspectives can be utilised to impact design and implementation of devices, we have shown that adopting a continual feedback loop when conducting HCI research is useful in ensuring that findings are accurate and have an appropriate level of nuance, but participant perspectives are also fickle and inconsistent to use as a central data source.

# Chapter 8

# Discussion

## 8.1 Overview

In this chapter, we provide a reflection upon the thesis as a whole and critically reflect upon it against our research questions, providing a narrative by taking a critical view towards our contributions and the role they play in our understanding of interaction in these contexts. To begin, we provide feedback and investigation in the form of interviewing 3 clinicians who took part in research across chapters 5 and 6 to understand the impact that participation had on them and the efficacy of our adopted methods. We then provide discussion in 3 dimensions, each taking the form of implications, before providing some general reflections upon the work and how it was conducted. We discuss design implications for devices in safety critical environments, both inside and beyond the diagnostic reporting office, the methodological implications of employing ethnographic techniques and their relevance to the topic, and the practice and pedagogical implications for users of these tools and devices in safety critical and medical situations. We conclude with critical reflections on the thesis as a whole, addressing our limitations and our own personal *lessons learned* from the research.

## 8.2 Interviewing Participants and Feedback

Several weeks after the last piece of research in Chapter 7 had been conducted, we approached participants of all research to take part in a 1 to 1 interview with the author to understand the impact that the research had on them - the purpose of this was

to gauge the effectiveness of our methods, and to explore if participating had actually resulted in changes to the way these practitioners interacted with the tools at their disposal.

3 participants responded, all of whom were consultants who had participated in Chapters 5 and 6. We conducted a semi-structured interview individually with each of them, taking place over teleconferencing software in a 1 on 1 capacity with just the author and the participant. Participants were asked a series of prompts (below) that the author would then expand upon and engage with. The audio was recorded and transcribed before being subjected to brief thematic analysis by the author inspired by the work of Braun and Clarke; an initial pass would be conducted where key phrases and quotes would be identified and tagged with commentary before a $2^{nd}$ pass of analysis was conducted with these potential themes in mind, codifying the initial codes into verified themes [55].

| **Clinician Questions** |
|---|
| Do You Remember The Purpose Of The Study That You Took Part In |
| Have You Taken Part In Any Other Outside Studies After That |
| Has Taking Part In The Research Made You Reconsider Any Aspects Of Your Day To Day Work Or The Way You Interact With Digital Systems |
| Do You Think The Research You Participated In Will Have A Positive Effect On The Future Of Your Work |
| Are You Interested In Participating In More Research Of This Kind |

Table 8.1: Clinician Interview Question Set

### 8.2.1 Results

Due to the fact that the sample size for our clinical participants was relatively small, we found the most appropriate way to reconstruct the interviews would be participant by participant, before summarising the common themes.

*P1*

P1 had been part of several qualitative studies regarding their interactions with devices both interventional and non-interventional. They gave a lengthy and all-encompassing answer that covered extensively the times were a participant, at times a co-analyst, and in their managerial position a facilitator of academic research, demonstrating an excellent

memory for all of the research they participated in, its purposes, and in some cases the outcomes such as presentations and peer reviewed papers.

They worked as a senior member of an educational facility, and as such felt a positive pre-disposition towards research, frequently working in interdisciplinary environments as a stakeholder of research projects, possibly influencing the high-quality recall and interest in outside research participation. They also opined their senior position at their organisation was a reason they participated in research as they would know the location and timings of a study, and felt that their presence as a participant would assist in influencing others at their organisation. As such, P1 provides a good example of acting as a *local champion*, an enthusiastic and well-respected member of their community that champions academic research altruistically for the benefit of the discipline. This was caveated with the fact that this was quite literally *local* championing; they believed that if the research held little relationship to their organisation or professional sphere, they would feel less motivation to take part, indicating that even with local champions there is a necessity to ensure professional fulfilment to improve engagement.

They spoke of enjoying participating in research for the long-term benefits it provided them and their coworkers in terms of furthering knowledge about their profession – they believed that this was the primary beneficial outcome of outsider academic research, the ability to utilise scientific and published reports to drive pedagogical improvement and systemic change in the ways in which their discipline was conducted as a whole. The caveat with this, however, is that they did not believe that participating in research had noticeable and significant effects on their personal conduct and working habits. In fact, they believed that the debrief and reading end publications further entrenched their behaviour, as they interpreted the results of the research they participated in as confirming that their methods were the most accurate and effective.

P1 provides us with an impact of the ideal participant-champion. A senior member of an organisation who can provide geographical, administrative and domain assistance as well as being a willing and enthusiastic participant in the research. However, it must be acknowledged that since P1 did not feel a personal benefit to their work by taking part, it was vital that the proposed research aligned with their organisation – the benefits of their

championing would only be accessible if the research was taking place on their "home ground".

*P2*

P2 had taken part in a laboratory-based study investigating interactions with a hardware device using modified Wizard of Oz software. They showed good signs of remembering the nature of the study, but could not recall the underlying purpose of the overall research, claiming

"*The information was there and I looked at it at the time... but I can't really remember the details, to be honest*"

This was the only academic study that P2 had participated in, but they claimed it had a significant and noticeable effect on the way in which they now use the piece of hardware that was the focus of the research. After discussing the results during the debrief with the research team, P2 realised they were an outlier compared to their peers in terms of interactional behaviour, and reflected on it:

"*One of the points [the researchers] said was that I seem to type a lot whereas other people would click on a word and dictate it... it just made me think about it. It made me re-evaluate a few things.*"

Following this reflection, P2 bought a new piece of hardware to allow for them to carry out everyday tasks in a way that did not significantly affect their existing workflow, but more closely aligned with the methods and preferences of their peers, something they found to improve their performance:

"*I wouldn't say it's changed the world for me, but since being involved in that research, I have found it to be beneficial, so I've stuck with [it]*"

When asked to reflect on their experiences as a *participant*, P2 claimed they would be open and enthusiastic towards taking part in further studies. This was partially down to the actionable and beneficial results they personally experienced, but also due to the

way in which the study was set up to suit them:

"*It was quick and easy wasn't it. I think if it was a massive time burden, I wouldn't have been so keen. It wasn't too much of a problem to participate in, but if it was a lot of time involvement that would have put me off*"

The study P2 took part in only required 30 minutes of their time, and simply required them to replicate their everyday working behaviour, meaning it was not cognitively complex or time consuming – the study was set up at the hospital site at which P2 worked, meaning they participated on their lunch break and were still able to complete their normal day's duties within the working hours. They did, however, highlight specifically that if these factors were not the case, they would have significantly impacted the likelihood of their participation:

"*I think if you said 'can you come to [location]' and spend a couple of hours there, then I would be less likely to take part, but they fact that they were all sort of convenient made it all easy*"

Indicating that this was not just a simple matter of preference, but a major factor that hinged their enthusiasm to take part in outside research.

P2 provides us with an example of a participant who genuinely felt that the research they took part in had a positive impact on the way in which they work, and enjoyed being a part of the study to the extent that they would be open to participating in the future, making them an invaluable contact and member of the medical community for ECRs. However, the external factors of the research being local, accessible and temporarily efficient play a huge role in initial stages of recruiting and conducting them in the study, demonstrating an important dimension to carrying out research with them in mind.

*P3*

P3 was an active participant and stakeholder in several academic studies, to the extent that they answered our opening prompt with

*"Sorry, I get a little confused because there's 2 PhD projects going on now"*

Referring to the fact that they have participated in studies relating to multiple distinct topics with significant deviations in their purpose and study design. Despite this, P3 was still able to give a thorough explanation of the relevant academic research that they had taken part in regarding the interview topic. They also spoke about being a facilitator of other academic studies alongside being a participant

*"I'm helping out [PhD student]... I haven't taken part in any of the research, but I have been supporting them in the projects they want to do; helping collect data and stuff like that basically, but not as part of the research"*

In terms of the impact that being a participant had on their working life, P3 was much like P1; they acknowledged that the research had a positive wider impact on the community, and retrospectively reflected that being a research participant was beneficial to their professional conduct as it encouraged critical evaluations on their behaviour and interactional patterns over just copying coworkers or following tacit intuitive patterns, but did not believe that they had any tangible practical benefits in terms of the quality of care they provided to patients as a result of taking part in the research.

The key factor that P3 discussed was their belief in being *"research positive"* and the impact that personal disposition had on the likelihood of sustained, quality engagement with academic studies. They opined that clinicians belonged to classifications of *willing to take part in research*, and *not willing to take part in research*, and that this was inherent – clinicians and medical professionals who fell into the latter category may still take part in research at the urging of a senior member of staff, but were unlikely to volunteer or remain engaged in longitudinal studies, in their belief. P3 claimed to be a part of the former group, a physician who was interested in research and enjoyed the idea of academia, and as such participated in studies willingly. Whilst we have no formal verification of this, it still has significant connotations for ECRs planning research projects and looking for their own champion if they have identified a medical domain expert is unwilling to take part in research – that it may be more valuable to their time in the long term to simply move on to a different, "*research positive*" member of the community.

In terms of motivation and likelihood of participation in more tangible terms, P3 followed suit of P2 – location, ease of access and the time required for them to take part were the key factors in dissuading them from participating in a research study. They enjoyed the idea of taking part in *"local"* research hosted by their employer or local organisation, but would be unwilling to travel significant distances or give up time that they would have otherwise spent at home outside of work hours to participate, regardless of the professional incentives to their work.

### 8.2.2 Reflections On The Thesis

After conducting these interviews, it becomes clear that there were some factors that may have been outside of our control in organising and conducting the most effective research; if what P3 claims is accurate and representative, we could have carried out a pre-study sift to only examine participants that were "*research positive*" to increase the quality of participation of our radiological cohorts. It is, however, equally possible that this sift was done for us by potential participants who simply did not volunteer to take part - it is difficult to gauge to what extent all participants across our studies were "*research positive*" but a wider survey of NHS radiology would merit interest to study this.

When projecting our findings and reflections back through these interviews, we do see some impact; P2 explicitly states that they have reconsidered and adjusted their interaction methods simply based on *participating* in research without even seeing the results of the analysis and the thesis - similarly, both P1 and P3 stated their optimism about participating in research that they believed to have a tangible and actionable outcome on the methods of practice and interaction. As such, we can gleam from these interviews that adopting this *work as imagined* approach allows us to meaningfully impact on the practices of participants as we are providing *observable* and *tangible* data that we can present back to them: by utilising their own in-the-wild interaction techniques that have come about through experience, we provide a point of comparison that participants (and readers who belong to the same domain) can leverage against our findings and recommendations.

Put simply, by taking our ethnomethodological and *work as done* lens and placing it on radiologist's methods of interactions, we allow participants to adjust and adapt based

on the recommendations we provide, as the data is purely taken from their current and observable habits and tendencies. This demonstrates the efficacy of our methods from the perspective of those who have actually participated in the thesis research.

## 8.3 Design Implications

Our first research question at the outset of this thesis was to investigate "*How Do Radiologists Interact with Safety Critical Devices In-The-Wild When Constructing Diagnostic Reports?*". Chapter 5 described the process of receiving, analysing, reporting and communicating all whilst stationed at the radiologist's desk. This, constructed from our non-interventional observations, provides an overview of the process that allows us to dive more in-depth into specific aspects of interaction that draw the eye. Chapter 5 explores the communication process, in both analogue and digital ways, with the myriad of systems at the reporter's disposal, and the way in which interaction is perceived to be entirely tacit and intuitive – this aspect reveals the importance that peer-to-peer teaching and folk theories have to practitioners when learning how to construct reports: the phrase "*it works better when I do that*" summarises this. Chapter 6 delved deeper into the interaction process by isolating the specific relationship between the radiologist and their communication vessel in the form of the Voice User Interface, examining the ways in which interactional techniques and a lack of transparency in design can cause a disconnect between the intentions of the practitioner and the end report. We also demonstrated the high variation between users even within a singular organisation. Our discussions with active practitioners about the devices they use and their efficacy in Chapters 4 and 7 also presented difficulty with untangling and separating design flaws from systemic ones. Here, we discuss these factors and how the augment our design recommendations found throughout the thesis.

### 8.3.1 Generic Design for Personalised Use

The implications for designers from these chapters are complicated. Whilst we did show specific ways in which existing devices are misused, misunderstood and potentially maligned, there is difficulty in offering designs that specifically tailor to the radiological community. This is because we frequently encountered high levels of idiosyncratic behaviour and heavily personalised techniques based upon the preferences, training and tacit knowledge of the hospital ecosystem on behalf of each individual user. When observing

radiologists in their natural working environment, factors such as the amount of time they have to communicate a result and the clinician they are attempting to reach may result in the same practitioner utilising different digital communication methods to deliver the same findings on a medical study. In the same way, offering a sample cohort of radiologists the same speech to text system and peripheral devices results in a high variation in successful transcription, allowed for by the "*intuitive*" nature of the devices - literature suggests that simply thinking of voice based interactive systems as *intuitive* is not enough, and here we see an affirmation of that concept [85, 252].

As such, there is the indication that (whilst they should still be developed with the *Work As Done* findings of our research in mind) these devices may benefit from their generic design, as it allows for the practitioner to imprint their personal perspectives and preferences of use onto them. Allowing for, and even advocating for, personalisation and variation in techniques of interaction would align with the radiological practices in other forms, such as allowing practitioners to write reports how they like as long as the medical opinion is accurate and actionable; Burns et al's work on structured reporting demonstrated that, regardless of the positive opinion of the recipient, forcing radiologists to construct structured and standardised reports had a negative impact on the quality of their output [60]. Additionally, as seen in the preceding MSc groundwork [80], trainee radiologists tend to view extant devices and systems through the lens of the technology and personal devices they use outside of the workplace – here, producing generic devices would lean further into this, allowing existing digital literacy gained outside of the workplace to inform interaction in the professional sphere [186].

The caveat here is of course in the *education* and *instruction* on these devices; allowing for personalisation may allow for alignment to other medical practices, but could also further entrench the bad habits we highlighted in Chapter 5 of *"It Works Better When I Do That"*. We will examine pedagogical implications further in this chapter, but it should be noted that the two should be closely related when offering design and implementation guidance to those in the medical device manufacturing community.

### 8.3.2 Political and Systemic Factors in Device Efficacy

Our design implications are further complicated by the difficulty in identifying and addressing individualised device and system points of failure from the end-user's perspectives. We have frequently discussed participant's tendency to discuss systemic and political problems during collaborative and participant-driven studies and the methodological caveats that it caused, but this also influences our evaluations and recommendations of existing systems. This is not the first evaluative study based around the NHS in modern times to face these issues [101, 51], but here it becomes apparent that the active practitioner's reckoning and evaluations of a device is interconnected to their general working conditions and perspectives on a site's organisation and cohesion. This further advocates the need for *holistic* understanding of the environment even beyond the distinct working environment, but also leads to the need to consider the context of implementation when evaluating device efficacy.

In simpler terms, the UK clinician's opinion of how suitable and usable a device is is tied to their opinion of their job situation as a whole. This means that our evaluations and design recommendations should also come with the acknowledgement that this study was conducted in the immediate Post COVID-19 Era, and all fieldwork was conducted under the political leadership of the Conservative Government. We do not wish to pass political judgement in this discussion, but it means that our holistic evaluations and the perspectives offered by participants on issues such as "manpower", "funding" and implementation and hierarchical complaints are of a particular time and place with reduced capacity in hospitals, low levels of funding for the NHS, and significant attrition and burnout rates from staff as a hangover of the pandemic era. An opinion piece by physician scientists Goldacre and MacKenna from 2020 is apt in summarising this position, advocating for a more fair and equal use of hospital data and funding [128] - the article situates our position as outside researchers in these studies, attempting to address longstanding and systemic arguments with small digital and interactional suggestions.

## 8.4 Methodological Implications

Our second key research question was *How Can the Perspectives of Diagnostic Practitioners Be Utilised to Meaningfully Impact the Design and Implementation of Devices In Safety Critical Environments?*. Chapters 4 and 7 specifically take a critical eye at the ways in which pure

participant perspectives can be utilised to generate research avenues, but this method was also employed heavily in the post data-gathering of Chapters 5 and 6. As such, we relied significantly on the radiologist as both a data source and as a reference, utilising their tacit knowledge and education as leverage against our own findings to add contextual and real-world nuances to what we present, but ensuring that our findings were continuously accurate and appropriate. Here, we examine the ways in which our chosen methods, and the ways in which we continued out "checking-in" approach affected the quality, quantity and genre of data we were able to collect.

### 8.4.1 Ethnomethodology, Collaboration and Efficacy

This thesis had aimed to answer a question based upon how practitioners interact with the tools at their disposal (the "practical action" [245]), and their personal perspectives on how this interactional process functions. Referring back to our *Approach* chapter, the goal was the "make visible the real-world socialty of a setting" [86]. As such, the thesis has primarily used *ethnomethodologically informed* methods of constructing an ethnographic record in order to study these methods and behaviours that constitute *use* and *interaction*. This has entailed conducting focus groups, interviewing participants individually, and non-interventional observations both in-the-wild and in laboratory conditions. By using these methods, we have been able to construct a multi-dimensional representation of how the practitioner interacts with the tools and devices at their disposal when constructing and communicating reports – we have been able to incorporate the user's perspectives and expertise alongside our own providing us with a level of tacit knowledge and understanding beyond vulgar competency (a term referring to the base level of knowledge a researcher needs to properly understand what they are observing [245], [183]), pushing our findings beyond surface-level observation into a more contextually aware reconstruction of interaction.

Whilst initially allowing participants to guide our decisions on where our focus of study should lie in Chapter 4, we adopted an *indifferent* approach to observing these patterns and activities in Chapters 5 and 6 [61]. This not only enabled us to utilise our expertise, but provided two representations of interaction with which we can compare and contrast *device failure* with *user error*. Examining these two representations reveals that participants had a tendency to overestimate the number of times "total failure" (i.e. a

complete shutdown of available systems or a crash that required full restart, preventing the user from completing their duties) occurred in a day-to-day setting. Our observations did not reveal any examples of such total failures, but instead revealed a high number of smaller, seemingly "invisible" errors that participants did not tend to highlight when directly asked, to the extent that they seemed unaware of these errors occurring when presented back with our comments in Chapter 5. This presence of "invisible errors" correlates strongly to existing literature on the ways in which failures compound in medical settings when design is unsuitable for purpose [41, 121, 122]. However, when presenting designs and proposals based on our own findings in Chapters 5 and 6, we found that participants did not perceive the problems we were seeking to solve as the most pressing issues.

Examining this journey of initialisation through to independent examination and back to feedback from our user base demonstrates both the effectiveness of our multi-dimensional approach, in that we were able to both validate and question the viewpoints of users and add our own in-depth insights into where areas of fault and friction occur, and the problems associated with choosing such a method – our user base did not always share our ideas of how these faults could be corrected, and would at times have difficulty aligning themselves to our research vision and the direction that participatory activities should take; it is clear that a fully participant-driven approach would further these problems. As such, the methods adopted in this thesis have their place *alongside* traditional formal methods as metrics of evaluating device use and safety.

### 8.4.2 Conducting Research as an "Outsider"

Here, we also wish to briefly discuss the unintended consequences of choosing to conduct research "in-the-wild" from a personal perspective, and offer reflections on the practicalities of collecting real world clinical data. This includes the mental and emotional burden of shadowing a clinicians day-to-day, as well as the paperwork considerations of wanting to conduct work in these spaces.

In Chapter 5, we chose to conduct our observational study in hospitals during peak times, situating ourselves in the radiology department as it was open to the public. Often, the finding of this radiology department caused some difficulty, as they were located behind locked doors or through different departments and wards of the hospital – this

required a great deal of trial-and-error, and often necessitated multiple instances of having to ask for help locating and accessing the reporting offices. As a student without any NHS identification, access often required clarification of the study purpose and validation from the practitioner we were there to shadow on that day. We frequently had to walk through active wards with patients, including paediatric wards, and seeing what the radiologist saw in terms of a medical study could on occasion mean viewing a serious injury or potentially distressing scan. Whilst we had been given a basic overview of the duties of the radiologist, seeing "the real thing" proved to produce a higher mental toil than was anticipated. Working out of these spaces as an unfamiliar and non-medical practitioner *thrown in at the deep end* (comment from P7 in Chapter 7) was difficult, and this had an impact on the data gathering process; it required some time prior to properly beginning the study to allow for us to calm, relax and enter a state befitting the task-at-hand, and rendered observing multiple sessions in a day inappropriate, as the quality of data gathered throughout the day would be inconsistent. Furniss briefly discusses a similar experience and how it had an emotional impact even as a seasoned and experienced academic, demonstrating how powerful an effect this environment can have on "*outsiders*" [120].

Finally, we wish to discuss the methodological choice and implications of utilising medical practitioners as the main source of data. A primary factor here, as mentioned in Chapters 2 through to 7, is the availability of clinicians. Wales, where our studies were all based, has the lowest *radiologists per 100,000 people* ratio of all of the 4 British nations, and it became clear early on in our research that high participant pools would be difficult to achieve [240]. This is, however, far from the only research to establish the difficulties in procuring an appropriate number of clinicians for in-person research. This is a well-established aspect of conducting studies in the domain [242, 113]. What was unexpected when conducting this research was the nature of the ways in which clinicians interacted with us when given the opportunity to contribute to research. Of particular note was the way we found there we became like *Agony Aunts* to our participants when we invited discussion and analysis on devices and workflow.

*Agony Aunt* (a reference to the advice columns in newspapers) refers to the frequency with which participants would shift conversations away from design and device commentary in order to instead discuss systemic and hierarchical problems that they faced within their

individual organisations and the NHS as a whole. This was especially present in Chapters 4 and 7, where groups of participants would often attempt move the topic of conversation in a way that expressed personal and professional frustrations with aspects that could not be addressed under the remit of the study in question. Examples of this included asking for a larger workforce and better financial support in Chapter 4, and discussing the problems with management and lack of top-down support when procuring appropriate systems in Chapter 7. Whilst these discussion points were useful in the ways they provided contextual information to us regarding aspects such as how systems are implemented, procured and maintained, internal frustrations with technology, and the overall state of NHS Wales as a semi-digitised healthcare organisation, they often detracted from the main goals of actionable outcomes. We found that utilising an entirely participant-led method of investigating design and implementation of tools in-situ would not be an appropriate method because of these factors, and as such adopted a mixed-methods approach throughout.

## 8.5 Practice and Pedagogical Implications

Beyond implications for the ways in which radiological devices are designed and implemented, it is worth considering the other side of interaction; the role of the user in the efficacy and efficiency of how clinical workflow operates in the reporting process. Throughout this thesis, we have continually included the participant in the research process both as a primary source of data and as a collaborative source of analysis and found that interaction is often not a codified and taught process, and variations in interaction could often lead to counts of device failure. The allowances for personalisation and preference in technique result in difficult oversight and can *muddy the waters* over issues in failure – we observed examples of users not operating buttons correctly, forgetting to trigger recordings on or off, or attempting to answer questions from colleagues whilst analysing a medical study. Further research efforts should be made on the best ways to interact with these devices in *practical* environments, and this work should then be translated into general conduct education.

For institutions such as the NIAW, the implications are clearer. Continually, practitioners expressed that they were not trained in elements with no direct medical underpinning – this included how best to interact with the digital devices at their disposal, typist work such as proofreading and editing, and even hospital workflow considerations such as the most efficient methods of communicating with a referring clinician when presented

with an actionable finding. Consistent reminders of best practice have been shown to be effective when attempting to change clinician's habits [107], and a study on RSI in radiologists found many were unfamiliar with the concept of HotKeys and other digital shortcuts, something rectified with a small training study and continual reminding of proper technique [95]. This is not solely the responsibility of the academic institution, but when presented with such factors it is evident that there is more to being a modern radiologist than medical training and this should be accommodated for.

## 8.6 Critical Reflections and Limitations

The key choices made in this thesis were made with a variety of factors in mind, all of which shaped the ways in which the research was carried out. When deciding upon research questions and the directions in which we would progress, it is impossible not to acknowledge the obligation we held to the National Data Resource and to the National Imaging Academy Wales to hold our focus on radiological tools. This meant that our focus was often held with providing actionable outcomes to the NIAW alongside our contributions to the HCI community. The nature of the collaboration is evident throughout our work, and as such, many avenues of research that could be explored in other safety critical environments are left open for future work to examine, and we had a narrow focus on NHS diagnostic medicine. This provided us with benefits in the support, recruitment and analysis of our results, but also meant that our overall agenda remained narrow throughout.

It is also worth acknowledging that we do little to investigate the individual construction and design of individual elements of hardware and software at the behest of examining "holistic" interaction. Whilst we offer insight into the ways in which, for example, the dictaphone and NUANCE software are interacted with, this thesis does not provide a deconstruction of the dictaphone itself, or an examination of the coding and metrics of the VUI software. Our emphasis was on examining how interaction was shaped by pre-existing devices with a view towards the holistic and humanistic elements of interaction, examining behaviour and social structure, but this means that we have not addressed formal metrics and verifications in place. We do not view this as undermining our findings, but acknowledge that this work goes hand-in-hand with evaluations of construction, ergonomics, electronics and programming. Our work addresses the design aspects of devices and software from the *Work As Done* perspective, and as such provides equal

guidance and advice to designers and manufacturers on how future devices will be used *in-situ* to allow for them to consider and accommodate for them. Further, we acknowledge that we do not offer physical designs of our own, only offering potential designs, mockups and facsimiles for evaluation by participants. Like with our acknowledgement above, we reinforce that this thesis focused on ethnographies of interaction with real-world tools – the primary goal was to examine how *Work Is Done* with the devices that practitioner would actually interact with when conducting their duties.

Finally, with a view towards the future, we believe that there is a likelihood that our insights will need to be taken in tandem with the growth of AI based systems in NHS hospitals. We have not addressed the implementation of AI particularly heavily in this thesis as it remains a controversial topic in its infancy, with many current implementations in their developmental stage and a distinct lack of codified guidance and oversight at the time of this thesis writing. The impact that AI and clinical decision support systems will have on the state of modern radiology remains to be seen, but it is worth considering that our holistic examinations of interaction are focused on *existing* tool use, and with digital medicine currently in flux with AI becoming increasingly hard to ignore, it is likely that these findings will need to be adapted to also consider how AI support tools will fit into diagnostic workflow and the hospital ecosystem.

# Chapter 9

# Conclusion

This thesis has focused primarily on the ways in which diagnostic practitioners interact with and perceive the digital devices and systems utilised when analysing medical imagery and constructing a diagnostic opinion. We have also conducted some analysis on the wider ways in which research of this style is carried out, with a view to improving the HCI and healthcare relationship and producing more sustained and high-quality actionable outcomes. Here, we will briefly revisit our problem description before examining our key findings from Chapters 4-7, before summarising our key contributions to the field and exploring where the future of study into this area may lie.

## 9.1 Chapter Summaries

### 9.1.1 Perspectives On Workflow From Clinical Practitioners

Here, we conduct a pair of focus groups with active practitioners examining their perspectives towards the state of clinical workflow and how it can be improved, with an emphasis on digital devices. The structure of the session began by asking participants to self-identify every stage of the patient interaction and treatment process, before asking them to identify the areas of fault and friction that occur across this timeline. We then asked participants to creatively discuss how digital implementations could be used to solve these problems.

Overall, 3 themes arose in the ways in which participants believed digitisation could be implemented to address areas of friction: Communication, Integration and Support and Infrastructure. Communication referred to improving the existing methods of messaging

and calling other clinicians in the hospital, and Integration referred to the perceived lack of interoperability and high volume of data and systems that participants had to interact with when completing tasks. Support and Infrastructure was a more abstract factor that referred to the systemic issues in the wider NHS Wales that participants believed hindered good procurement and implementation of high-quality reporting devices and systems.

This exercise helped to demonstrate the ways in which clinicians believed hierarchical and managerial issues played an equal part to the failures of digital systems, and illustrated an "ease of use" factor in what they desired from future systems that will become prevalent in later investigations.

### 9.1.2 Observations Of Communication and Interaction In-Situ

Following a participant-led focus group, we conducted a non-interventional observational study of communication and interaction in-the-wild. We observed 11 diagnostic practitioners including consultant radiologists and reporting radiographers for a total of 35 hours of data, collected under the 7 dimensions as laid out in Spradley's framework for ethnographic data gathering. These observations comprised of situating ourselves in the radiologist's reporting office and note-taking for a 3 hour "session", followed by follow up interviews with the participants to achieve a full understanding of behaviours and interaction patterns.

Alongside a reconstruction of the reporting process through thick description, we discussed the interesting and counter-intuitive ways in which practitioners interacted with their dictaphones, choosing to speak softly and monotonously to (in their beliefs) maximise the speed and efficiency with which the speech recognition software recognised their talking. We highlighted the variety and lack of codification surrounding the ways in which a result could be communicated, ranging from instant messaging to internal phonelines and in-person conversation, with emphasis being placed on both parties to the communication's preferences for how they liked to be reached. Finally, we drew attention to the heavily interrupted nature of attempting to focus on a report for extended periods of time, demonstrating that the levels of severity that interruption would require was unimportant, and that all interruptions were treated equally in terms of the attention demanded from the practitioner in focus.

From these findings, we can readily conclude that there is a noticeable lack of standardisation and codification with the interactional aspects of analysing and communicating when authoring reports; by this, we mean that radiological practice is focused heavily on ensuring medical accuracy and a high tempo of work, and the smaller and personalised aspects of what go into writing reports are left up to the choice of the practitioner. This leads to divergence in the ways in which tools are used, and means that there is a requirement for a variety to satisfy all parties' needs. This may, in part, necessitate a generic nature of design and implementation.

### 9.1.3 Investigating Voice User Interface Interactions When Constructing Reports

One of the key areas of intrigue identified in Chapter 5 was the seemingly counter-intuitive way in which practitioners interacted with their dictaphones and speech to text systems. The phrase "it works better when I do that" was frequently applied to justify techniques, but with little further explanation given. Accordingly, we chose to conduct a more thorough and controlled investigation into the ways in which practitioners utilised voice user interfaces when reporting. This took the form of a laboratory study where we asked consultant and trainee radiologists to report on real-world anonymised medical studies, with the session being recorded on a series of cameras for in-depth analysis. Following analysis of these sessions, we invited a portion of the participant pool to conduct collaborative discussion and analysis on their own data, to offer insight and contextual explanation on the origins and efficacy of their behaviours, by playing them fragments of their session and inviting commentary.

Alongside a more structured reconstruction of the interaction process than what was able to be achieved in Chapter 5, we demonstrated the ways in which errors are produced, through a mixture of user and system error, and the ways in which they are corrected both physically, using the digital devices available to the practitioner, and temporarily during various phases of the reporting period. We highlighted the lack of proper design implementation with regards to mistakes, drawing attention to system errors that result in a semantically correct end report, meaning they are less likely to be spotted in the proofing process, but misrepresent the author's diagnostic opinion.

These findings re-instated the lack of standardisation and codification identified in Chapter 5, but offer a more in-depth analysis of the radiologist's role as a transcriptionist and editor despite the lack of provided training in the discipline. We demonstrate gaps in design in terms of the lack of 2-way communication (something sought after by users) and the lack of appropriate highlighting of errors, and offer ways in which VUIs in this context can be more appropriately designed to suit the practical Work As Done nature of the modern radiologist.

### 9.1.4 Design Guidance and Feedback From Expert Users

Here, we critically reflect upon our findings in Chapters 4, 5 and 6 by offering clinical practitioners the opportunity to comment and discuss our design ideas based upon our own conclusions. We held a pair of focus groups with both diagnostic clinicians and technical radiological specialist managers from local NHS health boards and provided them with discussion prompts based on our findings, before having them evaluate design prompts and offer their own perspectives on what can be done to improve current standards of practice and how they feel future systems could be designed to solve the issues brought up in the sessions.

Our findings revealed a generally negative opinion between practitioners and technicians: the technical participants opined of our design prompts that radiologists "just won't use it", in reference a perceived stubbornness to adopt new elements into a streamlined and personalised workflow, whilst our radiological participants were overtly negative over the perceived standards of implementation and maintenance from technicians.

From our investigations, we found 2 specific themes that occurred throughout all Chapters but were displayed more plainly here: **Be Non-Invasive** and **Don't Slow Me Down**. These themes are prevalent from the ways in which both participant groups perceived the design prompts they were given, and how they would improve them – the ideas of co-pilots, hand's free systems and voice assistants were provided as ways participants would prefer novel systems to be implemented, demonstrating a desire for diagnostic clinicians to preserve their personalised workflow as much as is possible.

## 9.2 Contributions

### 9.2.1 An ethnomethodological reconstruction of how an NHS radiologist constructs a report in-situ

Designing, building and implementing devices in medical devices requires a more nuanced understanding of "safety" and appropriateness than factory metrics and independent analysis in isolation. A holistic, user-centred understanding of the wider ecosystem of interaction, communication and *Work as Done* of a safety critical environment allows for devices to be designed and implemented with a functioning and all-encompassing knowledge of how they will be used *in-situ*, increasing their functional safety, and the likelihood of sustained and enthusiastic adoption from end users.

This thesis presents a pair of ethnographies that provide a naturalistic and transferable vignette into the interactional process that constitutes authoring a radiology report. We provide an overview of the wider hospital ecosystem through our investigation and analysis of communication methodologies and interrupted activity, but also demonstrate a detailed, in-depth examination of how practitioners interact with the key tools at their disposal when constructing and conveying their diagnostic opinions. By employing collaborative and participatory analysis methods, we include the clinician's perspectives and opinions to provide a real-world and experienced contextual underpinning of these vignettes, meaning that the output we provide to designers of medical devices is informed by non-interactive and impartial research, supported and contextualised with tacit knowledge.

### 9.2.2 A clinical perspective on where elements of friction and failure exist when diagnosing and treating patients, and how digital and non-digital interventions should be designed and implemented to address them

Throughout this thesis, we have heavily utilised (and at times relied upon) the perspectives, experience and domain expertise of end-users when designing studies, establishing competency and verifying our findings. This has allowed us to ensure that we remained focused on the issues and tasks that clinicians felt genuinely affected their abilities to conduct care of the highest standards, and meant that when drawing conclusions and moving onto more in-depth investigations we remained in a close relationship to the medical community.

Working in this fashion had specific outcomes that provided genuine insights into an interdisciplinary mixed methods approach for future researchers. By conducting an initial investigation in Chapter 4, we arrived at several conclusions based on participant's perspectives: that technological methods of communication in medical environments were flawed, and caused delay and frustration; that the standards of the technological systems available to clinicians were not appropriate; and that the ecosystem of constructing and communicating reports. This led us to conduct an observational study, impartially examining those claims, and then a laboratory study in further detail. These investigations provided previously unreported aspects of the communicative and interactional practices of radiologists in the workplace, valuable insights as a result of beginning our research with collaborative exercises and allowing end-users to point us to where we should be examining.

Similarly, as part of these studies, including discussion on our findings either through interviews or through collaborative analysis allowed us to further contextualise our results; for example, understanding that radiologists are not taught how to properly communicate with their VUIs provides a more nuanced implication of why potential semantic and non-semantic errors happen, as opposed to simple device failure. However, these were not always directly actionable – we found that clinicians were not always accurate in their identification of areas of failure (such as in claiming a frequent crashing of computer systems), at times struggled to align their perspectives with the end-goals of the research (such the consistent tendency to shift conversation away from the implementations and adjustments of technological systems into systemic and political disagreements) and the dismissal of ideas presented back to them, even of their own origin (such as in Chapter 7, where the mock up designs came from clinical perspectives).

It is clear that working in this fashion of continuous clinician involvement in HCI research as the main anchor point allowed us deeper access, more support and led to outcomes and findings that are more directly translatable to real-world practice due to the influences of heavily including clinicians, but was far from a perfect methodology – as such, we present an effective way of working *with* clinicians to examine causes of friction that ensures a user-centred approach and allows for actionable findings, but also provide a nuanced discussion on the caveats of that method.

### 9.2.3 A series of design and pedagogical recommendations for those constructing, implementing and using devices in health-based safety critical environments

In all of our content chapters and our discussion, we included recommendations based upon our findings to the reader, the user, and the designer. These recommendations are built upon the previous 2 contributions of this thesis, and as such we go beyond simply offering our own reconstruction of the radiologist's practice and the according clinical perspective; we expand upon and critically analyse the existing safety critical context to provide *actionable* and *tangible* findings for those who work in the domain.

Aspects of this for designers such as understanding the nature of radiologist's interactions with their VUIs and the way they perceive the digital systems at their disposal can be effectively leveraged into future designs; we had demonstrated unique ways of interacting with microphone-based systems and the reasons that failures may be missed, as well as presenting design prompts and adjustments for existing systems that would constitute improving the current state of clinical workflow (or, alternatively, why new implementations may not be adopted). These recommendations come grounded in both HCI and clinical experience, and take their evidence from *Work As Done* to provide a unique but ultimately confident and comprehensive perspective.

We also demonstrate to the institutional readers of this work, those who write, codify, deliver and assess radiological practice, the importance of understanding this reconstruction of interaction. Throughout our research, we continually highlight the through-narrative that participants adopt methods through uncodified and untested means, often claiming *"it works better when I do that"* without any measure of feedback or structural evidence. We offer opportunities for learning, structuring and investigating these systemic methods (such as in Chapter 5 the multitude of methods for communicating an actionable finding, based entirely on the practitioner and their organic community) in order to better understand the most effective ways of treating a patient, the ultimate end-goal of this work.

## 9.3 Future Work

### 9.3.1 In-Situ Implementation Of Our Work

We have offered mock up designs in Chapters 6 and 7, and offered potential solutions to the areas of friction identified in the organisation and ecosystem of the reporting process, but it would be of merit to examine physical interpretations of these ideas implemented and studied longitudinally in-situ. Clinical participants continuously noted in Chapter 7 that an area they took issue with was the *implementation* of technology, and that devices and software did not intrinsically have fault. Therefore, a particular area of study that expands upon this thesis would be to produce a device that aligns with our given advice of what a good piece of radiologist-facing should conform to, and conduct a mock procurement and implementation study that both verifies clinical perspectives that implementation strategies are inherently flawed and identifies the reasons behind these flaws, and then works to address them.

In short, what this thesis has done is present to designers of medically-focused and safety critical devices what we believe the areas of fault and friction are in the interactional process of authoring a radiology report, a highly complex task performed in a high pressure and cognitively noisy environment, and suggested ways in which they could be addressed - further work should seek to act upon these findings.

### 9.3.2 Expanding Horizons

The majority of this thesis has placed a specific lens on radiology. We have provided explanations and motivations towards this focus, but it is worth considering that it is not the only discipline that works in this way. Cardiology, pathology, genomics and other medical areas have practitioners that work independently on digital image analysis. Despite frequent allusions to the "uniqueness" of radiology both from our own conclusions and self-reported from participants, these disciplines would likely have a significant overlap with radiology, and would therefore value the takeaways provided in this research, but it is also likely that they converge significantly in terms of training and sociological interactions within the hospital. Comparing ethnographic in-the-wild studies is common practice, and here we feel it would be a worthwhile endeavour to examine the juxtapositions that would occur from replicating these studies. Beyond this, we have offered insight into interactions

that have merit being applied in other safety critical environments; literature has revealed the myriad of ways that other control rooms and organisational centres have failings that can be addressed through ethnographic study, and our findings should be extrapolated to situations that require rapid, accurate and conclusive reporting on expert opinions.

## 9.4 Concluding Remarks

In summary, this thesis has conducted a collection of ethnomethodologically informed ethnographies into the ways in which radiologists in the NHS use technology when authoring reports. We incorporate the clinician's own perspectives and experiences when drawing conclusions, and frequently involve the user as a collaborator when analysing data, providing us with a multidimensional and nuanced reconstruction of the habits, behaviours, preferences and patterns that go into the process of "reporting on a patient", whilst providing commentary and reflections on how this research is and can be conducted.

This work holds value to a series of stakeholders, from designers of tools and devices in medical contexts, educators and mentors in the diagnostic community, procurement and implementation managers, to radiologists themselves to reflect on interactional decisions and best practices. It enables for the more appropriate and high level construction of tools and devices in this area, and there is hope it can improve the standards of care that can be provided to patients in need.

# Bibliography

[1] Aaron C. Abajian and Hoiwan Cheung. 2022-07. No Knew Typos: Modernizing Radiology Dictation. *Radiology: Artificial Intelligence* 4, 4 (2022-07), e220105. DOI: http://dx.doi.org/10.1148/ryai.220105 Publisher: Radiological Society of North America.

[2] Joanna Abraham and Madhu C. Reddy. 2008-11-08. Moving patients around: a field study of coordination between clinical and non-clinical staff in hospitals. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW '08)*. Association for Computing Machinery, 225–228. DOI:http://dx.doi.org/10.1145/1460563.1460598

[3] Carlos Monroy Aceves, Patrick Oladimeji, Harold Thimbleby, and Paul Lee. 2013-07. Are prescribed infusions running as intended? Quantitative analysis of log files from infusion pumps used in a large acute NHS hospital. *British Journal of Nursing* 22 (2013-07), 15–21. Issue Sup13. DOI:http://dx.doi.org/10.12968/bjon.2013.22.Sup13.15 Publisher: Mark Allen Group.

[4] Faisal Ahmed, Yevgen Borodin, Yury Puzis, and I. V. Ramakrishnan. 2012-04-16. Why read if you can skim: towards enabling faster screen reading. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*. ACM, 1–10. DOI: http://dx.doi.org/10.1145/2207016.2207052

[5] Zaboor Ahmed, Khurram Khaliq Bhinder, Amna Tariq, Muhammad Junaid Tahir, Qasim Mehmood, Muhammad Saad Tabassum, Muna Malik, Sana Aslam, Muhammad Sohaib Asghar, and Zohaib Yousaf. 2022-04. Knowledge, attitude, and practice of artificial intelligence among doctors and medical students in Pakistan:

A cross-sectional online survey. *Annals of Medicine & Surgery* 76 (2022-04). DOI: http://dx.doi.org/10.1016/j.amsu.2022.103493

[6] Ahmad Al-Aiad, Ahmad K. Momani, Yazan Alnsour, and Mohammad Alsharo. 2020-08-10. The Impact of Speech Recognition Systems on The Productivity and The Workflow in Radiology Departments: A Systematic Review. *AMCIS 2020 TREOs* (2020-08-10). https://aisel.aisnet.org/treos_amcis2020/62

[7] Morana Alac. 2008-08. Working with Brain Scans: Digital Images and Gestural Interaction in fMRI Laboratory. *Social Studies of Science* 38, 4 (2008-08), 483–508. DOI: http://dx.doi.org/10.1177/0306312708089715

[8] Erik M Altmann and J Gregory Trafton. 2004. Task Interruption: Resumption Lag and the Role of Cues. *Proceedings of the 26th Annual Conference of the Cognitive Science Soceity* (2004).

[9] Erik M Altmann and J Gregory Trafton. 2007. Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin and Review* 14.6 (2007).

[10] Kirk St. Amant. 2017. The cultural context of care in international communication design: a heuristic for addressing usability in international health and medical communication. *Commun. Des. Q. Rev* 5, 2 (Aug. 2017), 62–70. DOI:http://dx.doi.org/10.1145/3131201.3131207

[11] Aloha Hufana Ambe, Margot Brereton, Alessandro Soro, Laurie Buys, and Paul Roe. 2019-05-02. The Adventures of Older Authors: Exploring Futures through Co-Design Fictions. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 1–16. DOI:http://dx.doi.org/10.1145/3290605.3300588

[12] Tariq Osman Andersen, Francisco Nunes, Lauren Wilcox, Enrico Coiera, and Yvonne Rogers. 2023-04-30. Introduction to the Special Issue on Human-Centred AI in Healthcare: Challenges Appearing in the Wild. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023-04-30), 1–12. DOI:http://dx.doi.org/10.1145/3589961

[13] Alyssa E Andrews, Raj M Ratwani, and J Gregory Trafton. 2009. Recovering from Interruptions: Does Alert Type Matter? *Proceedings Of the Human Factors and Ergonomics Society* 53.4 (2009).

[14] G. Antoniol, R. Fiutem, R. Flor, and G. Lazzari. 1993. Radiological reporting based on voice recognition. In *Human-Computer Interaction (Lecture Notes in Computer Science)*, Leonard J. Bass, Juri Gornostaev, and Claus Unger (Eds.). Springer, 242–253. DOI: http://dx.doi.org/10.1007/3-540-57433-6_53

[15] Mohammad R. Arbabshirani, Brandon K. Fornwalt, Gino J. Mongelluzzo, Jonathan D. Suever, Brandon D. Geise, Aalpen A. Patel, and Gregory J. Moore. 2018-12. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *npj Digital Medicine* 1, 1 (2018-12), 9. DOI:http://dx.doi.org/10.1038/s41746-017-0015-z

[16] Barun Aryal, Derek A. Khorsand, and Theodore J. Dubinsky. 2018-09-01. The Clinical and Medicolegal Implications of Radiology Results Communication. *Current Problems in Diagnostic Radiology* 47, 5 (2018-09-01), 287–289. DOI:http://dx.doi.org/10.1067/j.cpradiol.2017.09.009

[17] Joan S. Ash, Marc Berg, and Enrico Coiera. 2004-03-01. Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors. *Journal of the American Medical Informatics Association* 11, 2 (2004-03-01), 104–112. DOI:http://dx.doi.org/10.1197/jamia.M1471

[18] Ahmed Ashour, Darren M. Ashcroft, and Denham L. Phipps. 2021-05-01. Mind the gap: Examining work-as-imagined and work-as-done when dispensing medication in the community pharmacy setting. *Applied Ergonomics* 93 (2021-05-01), 103372. DOI: http://dx.doi.org/10.1016/j.apergo.2021.103372

[19] Mirjam Augstein, Thomas Neumayr, and Sebastian Pimminger. 2019. WeldVUI: Establishing Speech-Based Interfaces in Industrial Applications. In *Human-Computer Interaction – INTERACT 2019 (Lecture Notes in Computer Science)*, David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris (Eds.). Springer International Publishing, 679–698. DOI:http://dx.doi.org/10.1007/978-3-030-29387-1_40

[20] Steven J. Baccei, Cole DiRoberto, Janet Greene, and Max P. Rosen. 2019-01-05. Improving Communication of Actionable Findings in Radiology Imaging Studies and Procedures Using an EMR-Independent System. *Journal of Medical Systems* 43, 2 (2019-01-05), 30. DOI:http://dx.doi.org/10.1007/s10916-018-1150-z

[21] Helen M. Baker. 1997-06. Rules Outside The Rules for Administration of Medication: A Study in New South Wales, Australia. *Image: the Journal of Nursing Scholarship* 29, 2 (1997-06), 155–158. DOI:http://dx.doi.org/10.1111/j.1547-5069.1997.tb01549.x

[22] Aruna D. Balakrishnan, Sara Kiesler, Jonathon N. Cummings, and Reza Zadeh. 2011-03-19. Research team integration: what it is and why it matters. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 523–532. DOI: http://dx.doi.org/10.1145/1958824.1958905

[23] Carina Banziger, Kirsty McNeil, Hui Lu Goh, Samantha Choi, and Ian A Zealley. 2023-05-01. Simple changes to the reporting environment produce a large reduction in the frequency of interruptions to the reporting radiologist: an observational study. *Acta Radiologica* 64, 5 (2023-05-01), 1873–1879. DOI:http://dx.doi.org/10.1177/02841851221139624 Publisher: SAGE Publications.

[24] P. Barach. 2000-03-18. Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. *BMJ* 320, 7237 (2000-03-18), 759–763. DOI: http://dx.doi.org/10.1136/bmj.320.7237.759

[25] Margot Barry, Wietske Kuijer, Anke Persoon, Loek Nieuwenhuis, and Nynke Scherpbier. 2021-04-07. Enabling visibility of the clinician-scientists' knowledge broker role: a participatory design research in the Dutch nursing-home sector. *Health Research Policy and Systems* 19, 1 (2021-04-07), 61. DOI:http://dx.doi.org/10.1186/s12961-021-00715-z

[26] Sarah Basma, Bridgette Lord, Lindsay M. Jacks, Mohamed Rizk, and Anabel M. Scaranelo. 2011-10. Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription. *AJR. American journal of roentgenology* 197, 4 (2011-10), 923–927. DOI:http://dx.doi.org/10.2214/AJR.11.6691

[27] Eric Baumer, June Ahn, Mei Bie, Elizabeth Bonsignore, Ahmet Börütecene, Oğuz 'Oz Buruk, Tamara Clegg, Allison Druin, Florian Echtler, Daniel Gruen, Mona Guha, Chelsea Hordatt, Antonio Krüger, Shachar Maidenbaum, Meethu Malu, Brenna Mcnally, Michael Muller, Leyla Norooz, Juliet Norton, and Jason Yip. 2014-04-26. CHI 2039: Speculative Research Visions. *Conference on Human Factors in Computing Systems - Proceedings* (2014-04-26). DOI:http://dx.doi.org/10.1145/2559206.2578864

[28] Eric P.S. Baumer, Timothy Berrill, Sarah C. Botwinick, Jonathan L. Gonzales, Kevin Ho, Allison Kundrik, Luke Kwon, Tim LaRowe, Chanh P. Nguyen, Fredy Ramirez, Peter Schaedler, William Ulrich, Amber Wallace, Yuchen Wan, and Benjamin Weinfeld. 2018-01-07. What Would *You* Do?: Design Fiction and Ethics. In *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. ACM, 244–256. DOI:http://dx.doi.org/10.1145/3148330.3149405

[29] Eric P.S. Baumer, Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik, and Geri Gay. 2014-06-21. Reviewing reflection: on the use of reflection in interactive system design. In *Proceedings of the 2014 conference on Designing interactive systems (DIS '14)*. Association for Computing Machinery, 93–102. DOI: http://dx.doi.org/10.1145/2598510.2598598

[30] Eric P. S. Baumer, Mark Blythe, and Theresa Jean Tanenbaum. 2020-07-03. Evaluating Design Fiction: The Right Tool for the Job. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. ACM, 1901–1913. DOI:http://dx.doi.org/10.1145/3357236.3395464

[31] David Bawden and others. 2008. Origins and concepts of digital literacy. *Digital literacies: Concepts, policies and practices* 30, 2008 (2008), 17–32.

[32] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M. Vardoulakis. 2020-04-23. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, 1–12. DOI: http://dx.doi.org/10.1145/3313831.3376718

[33] Jessica Bell, Mirko Ancillotti, Victoria Coathup, Sarah Coy, Tessel Rigter, Travis Tatum, Jasjote Grewal, Faruk Berat Akcesme, Jovana Brkić, Anida Causevic-Ramosevac, Goran Milovanovic, Marianna Nobile, Cristiana Pavlidis, Teresa Finlay, Jane Kaye, and ELSI2.0. 2016-07-08. Challenges and opportunities for ELSI early career researchers. *BMC Medical Ethics* 17, 1 (2016-07-08), 37. DOI: http://dx.doi.org/10.1186/s12910-016-0121-5

[34] L. T. O. Bell, R. James, J. A. Rosa, A. Pollentine, G. Pettet, and P. McCoubrie. 2018-08-01. Reducing interruptions during duty radiology shifts, assessment of its benefits and

review of factors affecting the radiology working environment. *Clinical Radiology* 73, 8 (2018-08-01), 759.e19–759.e25. DOI:http://dx.doi.org/10.1016/j.crad.2018.04.007

[35] Gal Ben-Yehudah and Yoram Eshet-Alkalai. 2018-04. The contribution of text-highlighting to comprehension: A comparison of print and digital reading. *Journal of Educational Multimedia and Hypermedia* 27, 2 (2018-04), 153–178. Publisher: Association for the Advancement of Computing in Education (AACE).

[36] R. Bentley, J. A. Hughes, D. Randall, T. Rodden, P. Sawyer, D. Shapiro, and I. Sommerville. 1992. Ethnographically-informed systems design for air traffic control. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work - CSCW '92*. ACM Press, 123–129. DOI:http://dx.doi.org/10.1145/143457.143470

[37] Suzanne V. Blackley, Jessica Huynh, Liqin Wang, Zfania Korach, and Li Zhou. 2019-04-01. Speech recognition for clinical documentation from 1990 to 2018: a systematic review. *Journal of the American Medical Informatics Association: JAMIA* 26, 4 (2019-04-01), 324–338. DOI:http://dx.doi.org/10.1093/jamia/ocy179

[38] Carolyn Blair, Paul Best, Patricia Burns, Anne Campbell, Gavin Davidson, Joe Duffy, Anne Johnston, Berni Kelly, campbell Killick, Denise Dermott, Alan Maddock, Claire McCartan, Paula Mcfadden, Anne McGlade, Lorna Montgomery, Sonia Patton, Dirk Schubotz, Brian Taylor, Fiona Templeton, and Jade Yap. 2022-05-16. 'Getting involved in research': a co-created, co-delivered and co-analysed course for those with lived experience of health and social care services. *Research Involvement and Engagement* 8 (2022-05-16). DOI:http://dx.doi.org/10.1186/s40900-022-00353-x

[39] Ann Blandford. 2018-12-21. Lessons from working with researchers and practitioners in healthcare. *Interactions* 26, 1 (2018-12-21), 72–75. DOI:http://dx.doi.org/10.1145/3292023

[40] Ann Blandford, Erik Berndt, Ken Catchpole, Dominic Furniss, Astrid Mayer, Helena Mentis, Aisling Ann O'Kane, Tom Owen, Atish Rajkomar, and Rebecca Randell. 2015-11-01. Strategies for conducting situated studies of technology use in hospitals. *Cognition, Technology & Work* 17, 4 (2015-11-01), 489–502. DOI:http://dx.doi.org/10.1007/s10111-014-0318-7

[41] Ann Blandford, Dominic Furniss, George Buchanan, Harold Thimbleby, and Paul Curzon. 2010. Whos looking Invisible problems with interactive medical devices *(WISH)*.

[42] Ann Blandford, Dominic Furniss, and Chris Vincent. 2014-09. Patient safety and interactive medical devices: Realigning work as imagined and work as done. *Clinical Risk* 20, 5 (2014-09), 107–110. DOI:http://dx.doi.org/10.1177/1356262214556550

[43] Daniel J. Blezek, Lonny Olson-Williams, Andrew Missert, and Pangiotis Korfiatis. 2021-12. AI Integration in the Clinical Workflow. *Journal of Digital Imaging* 34, 6 (2021-12), 1435–1446. DOI:http://dx.doi.org/10.1007/s10278-021-00525-3

[44] Deborah A. Boehm-Davis and Roger Remington. 2009. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention* 41, 5 (2009), 1124–1129. DOI: http://dx.doi.org/https://doi.org/10.1016/j.aap.2009.06.029

[45] Susanne Boll, Philippe Palanque, Alexander G. Mirnig, Jessica Cauchard, Margareta Holtensdotter Lützhöft, and Michael S. Feary. 2020-04-25. Designing Safety Critical Interactions: Hunting Down Human Error. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, 1–7. DOI:http://dx.doi.org/10.1145/3334480.3375148

[46] Albert Boonstra, Arie Versluis, and Janita F. J. Vos. 2014-09-04. Implementing electronic health records in hospitals: a systematic literature review. *BMC Health Services Research* 14, 1 (2014-09-04), 370. DOI:http://dx.doi.org/10.1186/1472-6963-14-370

[47] Paul A. Booth. 1991-12-01. Errors and theory in human-computer interaction. *Acta Psychologica* 78, 1 (1991-12-01), 69–96. DOI:http://dx.doi.org/10.1016/0001-6918(91)90005-K

[48] Shadan Sadeghian Borojeni, Abdallah El Ali, Wilko Heuten, and Susanne Boll. 2016-10-23. Peripheral Light Cues for In-Vehicle Task Resumption. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. Association for Computing Machinery, 1–4. DOI:http://dx.doi.org/10.1145/2971485.2971498

[49] David Borys. 2009. EXPLORING RISK-AWARENESS AS A CULTURAL APPROACH TO SAFETY: EXPOSING THE GAP BETWEEN WORK AS IMAGINED AND WORK AS ACTUALLY PERFORMED. *Safety Science Monitor* 13 (2009).

[50] Jonathan Bowen. 2000-04-01. The ethics of safety-critical systems. Commun. ACM 43, 4 (2000-04-01), 91–97. DOI:http://dx.doi.org/10.1145/332051.332078

[51] Simon Bowen, Kerry McSeveny, Eleanor Lockley, Daniel Wolstenholme, Mark Cobb, and Andy Dearden. 2013-12-01. How was it for you? Experiences of participatory design in the UK health service. *CoDesign* 9, 4 (2013-12-01), 230–246. DOI:http://dx.doi.org/10.1080/15710882.2013.846384 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/15710882.2013.846384.

[52] Adrian P. Brady, Regina G. Beets-Tan, Boris Brkljačić, Carlo Catalano, Andrea Rockall, Michael Fuchsjäger, and European Society of Radiology (ESR). 2022-06-04. The role of radiologist in the changing world of healthcare: a White Paper of the European Society of Radiology (ESR). *Insights into Imaging* 13, 1 (2022-06-04), 100. DOI:http://dx.doi.org/10.1186/s13244-022-01241-4

[53] Jeffrey Braithwaite, Robert L. Wears, and Erik Hollnagel. 2016-10-03. *Resilient Health Care, Volume 3: Reconciling Work-as-Imagined and Work-as-Done*. CRC Press. Google-Books-ID: MDQNDgAAQBAJ.

[54] Lee Brasseur. 2012-01-01. Sonographers' Complex Communication during the Obstetric Sonogram Exam: An Interview Study. *Journal of Technical Writing and Communication* 42, 1 (2012-01-01), 3–19. DOI:http://dx.doi.org/10.2190/TW.42.1.b Publisher: SAGE Publications Inc.

[55] Virginia Braun and Victoria Clarke. 2006-01-01. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006-01-01), 77–101. DOI:http://dx.doi.org/10.1191/1478088706qp063oa Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa.

[56] Rosanna L. Breen. 2006-11. A Practical Guide to Focus-Group Research. *Journal of Geography in Higher Education* 30, 3 (2006-11), 463–475. DOI:http://dx.doi.org/10.1080/03098260600927575

[57] Mindaugas Briedis. 2019-10. Phenomenological ethnography can lead to the improvement of radiology diagnostics. *Adaptive Behavior* 27, 5 (2019-10), 347–350. DOI: http://dx.doi.org/10.1177/1059712319861663

[58] Mindaugas Briedis. 2020-04. Phenomenological ethnography of radiology: expert performance in enacting diagnostic cognition. *Phenomenology and the Cognitive Sciences* 19, 2 (2020-04), 373–404. DOI:http://dx.doi.org/10.1007/s11097-019-09612-x

[59] Nico Buls, Nina Watté, Koenraad Nieboer, Bart Ilsen, and Johan de Mey. 2021-03-01. Performance of an artificial intelligence tool with real-time clinical workflow integration – Detection of intracranial hemorrhage and pulmonary embolism. *Physica Medica: European Journal of Medical Physics* 83 (2021-03-01), 154–160. DOI:http://dx.doi.org/10.1016/j.ejmp.2021.03.015 Publisher: Elsevier.

[60] Judah Burns, Tara M. Catanzano, Pamela W. Schaefer, Vikas Agarwal, Derek Kim, Reece J. Goiffon, and Sheryl G. Jordan. 2022-05-01. Structured Reports and Radiology Residents: Friends or Foes? *Academic Radiology* 29 (2022-05-01), S43–S47. DOI: http://dx.doi.org/10.1016/j.acra.2020.10.014

[61] Graham Button, Michael Lynch, and W. W. Sharrock. 2023. *Ethnomethodology, conversation analysis and constructive analysis: on formal structures of practical action*. Routledge.

[62] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021-04-22. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5 (2021-04-22), 188:1–188:21. Issue CSCW1. DOI:http://dx.doi.org/10.1145/3449287

[63] Fredrik Bååthe. 2015-11-20. *Physicians' engagement: qualitative studies exploring physicians' experiences of engaging in improving clinical services and processes*. https://gupea.ub.gu.se/handle/2077/40438 Accepted: 2015-11-20T11:42:35Z.

[64] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019-11-07. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 3 (2019-11-07), 104:1–104:24. Issue CSCW. DOI:http://dx.doi.org/10.1145/3359206

[65] Alessandro Cannavale, Mariangela Santoni, Paola Mancarella, Roberto Passariello, and Paolo Arbarello. 2013-04-03. Malpractice in Radiology: What Should You

Worry About? *Radiology Research and Practice* 2013 (2013-04-03), e219259. DOI: http://dx.doi.org/10.1155/2013/219259 Publisher: Hindawi.

[66] Cheri L. Canon, Jeffrey Forris Beecham Chick, Ivan DeQuesada, Richard B. Gunderman, Noelle Hoven, and Ashley Elizabeth Prosper. 2022-02. Physician Burnout in Radiology: Perspectives From the Field. *American Journal of Roentgenology* 218, 2 (2022-02), 370–374. DOI:http://dx.doi.org/10.2214/AJR.21.26756 Publisher: American Roentgen Ray Society.

[67] Vin Capasso and Michelle Johnson. 2012. Improving the Medicine Administration Process by Reducing Interruptions. *Journal of Healthcare Management* 57, 6 (2012), 384.

[68] Paulo V.R. Carvalho, Isaac L. Dos Santos, and Mario C.R. Vidal. 2006-03. Safety implications of cultural and cognitive issues in nuclear power plant operation. *Applied Ergonomics* 37, 2 (2006-03), 211–223. DOI:http://dx.doi.org/10.1016/j.apergo.2005.03.004

[69] Paulo V. R. De Carvalho. 2006. Ergonomic field studies in a nuclear power plant control room. *Progress in Nuclear Energy* 48, 1 (2006), 51–69. DOI:http://dx.doi.org/https://doi.org/10.1016/j.pnucene.2005.04.001

[70] Paulo Victor Rodrigues De Carvalho, José Orlando Gomes, Gilbert Jacob Huber, and Mario Cesar Vidal. 2009-05. Normal people working in normal organizations with normal equipment: System safety and cognition in a mid-air collision. *Applied Ergonomics* 40, 3 (2009-05), 325–340. DOI:http://dx.doi.org/10.1016/j.apergo.2008.11.013

[71] Paulo Victor R. de Carvalho, Angela W. Righi, Gilbert J. Huber, Caio de F. Lemos, Alessandro Jatoba, and José Orlando Gomes. 2018-04-01. Reflections on work as done (WAD) and work as imagined (WAI) in an emergency response organization: A study on firefighters training exercises. *Applied Ergonomics* 68 (2018-04-01), 28–41. DOI: http://dx.doi.org/10.1016/j.apergo.2017.10.016

[72] Inha Cha, Ajit G. Pillai, and Richmond Y. Wong. 2024-07-01. Ethics Pathways: A Design Activity for Reflecting on Ethics Engagement in HCI Research. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, 3515–3533. DOI:http://dx.doi.org/10.1145/3643834.3660714

[73] Stephen Chan and Eliot L Siegel. 2019-02. Will machine learning end the viability of radiology as a thriving medical specialty? *The British Journal of Radiology* 92, 1094 (2019-02), 20180416. DOI:http://dx.doi.org/10.1259/bjr.20180416

[74] Chian A. Chang, Rodney Strahan, and Damien Jolley. 2011-08-01. Non-Clinical Errors Using Voice Recognition Dictation Software for Radiology Reports: A Retrospective Audit. *Journal of Digital Imaging* 24, 4 (2011-08-01), 724–728. DOI:http://dx.doi.org/10.1007/s10278-010-9344-z

[75] Minh Chau. 2020-12-01. Cultural Diversity and the Importance of Communication, Cultural Competence, and Uncertainty in Radiography. *Journal of Medical Imaging and Radiation Sciences* 51, 4 (2020-12-01), S17–S22. DOI:http://dx.doi.org/10.1016/j.jmir.2020.04.005

[76] EunJeong Cheon and Norman Makoto Su. 2018-02-26. Futuristic Autobiographies: Weaving Participant Narratives to Elicit Values around Robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 388–397. DOI:http://dx.doi.org/10.1145/3171221.3171244

[77] George Chin, Mary Beth Rosson, and John M. Carroll. 1997-03-27. Participatory analysis: shared development of requirements from scenarios. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. ACM, 162–169. DOI:http://dx.doi.org/10.1145/258549.258654

[78] Katie Chockley and Ezekiel Emanuel. 2016-12-01. The End of Radiology? Three Threats to the Future Practice of Radiology. *Journal of the American College of Radiology* 13, 12 (2016-12-01), 1415–1420. DOI:http://dx.doi.org/10.1016/j.jacr.2016.07.010

[79] Y. K. E. Choong, J. Tay, A. N. Y. Binte Abdul Razak, M. A. Look, and L. H. Sng. 2021-08-01. Bridging the communication gap between radiographers and patients to improve chest radiography image acquisition: A multilingual solution in the COVID-19 pandemic. *Radiography* 27, 3 (2021-08-01), 873–876. DOI:http://dx.doi.org/10.1016/j.radi.2021.02.004

[80] Rory Clark, Martin Porcheron, Matt Jones, Phillip Wardle, and Victoria Whitchurch. 2022. Perspectives On Machine Learning and Artificial Intelligence from Trainee Radiologists. *European Congress of Radiology* (2022), 911 words. DOI:http://dx.doi.

org/10.26044/ECR2022/C-21806 Artwork Size: 911 words Medium: text/html Publisher: European Congress of Radiology.

[81] Rory Stuart Clark, Tom Owen, Matt Jones, Martin Porcheron, and Phillip Wardle. 2023-10-26. It Works Better When I Do That: Interaction and Communication In Radiology Departments. In *Proceedings of the 41st ACM International Conference on Design of Communication*. ACM, 55–62. DOI:http://dx.doi.org/10.1145/3615335.3623011

[82] P. J. Clarkson, P. Buckle, R. Coleman, D. Stubbs, J. Ward, J. Jarrett, R. Lane, and J. Bound. 2004-04-01. Design for patient safety: A review of the effectiveness of design in the UK health service. *Journal of Engineering Design* 15, 2 (2004-04-01), 123–140. DOI:http://dx.doi.org/10.1080/09544820310001617711 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09544820310001617711.

[83] J. M. Coble, J. S. Maffitt, M. J. Orland, and M. G. Kahn. 1995. Contextual inquiry: discovering physicians' true needs. *Proceedings of the Annual Symposium on Computer Application in Medical Care* (1995), 469–473. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2579137/

[84] E. Coiera and V. Tombs. 1998-02-28. Communication behaviours in a hospital setting: an observational study. *BMJ (Clinical research ed.)* 316, 7132 (1998-02-28), 673–676. DOI:http://dx.doi.org/10.1136/bmj.316.7132.673

[85] Eric Corbett and Astrid Weber. 2016-09-06. What can I say? addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. Association for Computing Machinery, 72–82. DOI: http://dx.doi.org/10.1145/2935334.2935386

[86] Andy Crabtree, David M. Nichols, Jon O'Brien, Mark Rouncefield, and Michael B. Twidale. 2000. Ethnomethodologically informed ethnography and information system design. *Journal of the American Society for Information Science* 51, 7 (2000), 666–682. DOI: http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:7<666::AID-ASI8>3.0.CO;2-5

[87] Andrew Crabtree, Tom Rodden, Peter Tolmie, and Graham Button. 2009-04-04. Ethnography considered harmful. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, 879–888. DOI:http://dx.doi.org/10.1145/1518701.1518835

[88] Andy Crabtree, Peter Tolmie, and Mark Rouncefield. 2013. "How Many Bloody Examples Do You Want?" Fieldwork and Generalisation. In *ECSCW 2013: Proceedings of the 13th European Conference on Computer Supported Cooperative Work, 21-25 September 2013, Paphos, Cyprus*, Olav W. Bertelsen, Luigina Ciolfi, Maria Antonietta Grasso, and George Angelos Papadopoulos (Eds.). Springer, 1–20. DOI:http://dx.doi.org/10.1007/978-1-4471-5346-7_1

[89] Alessia Cristiano, Stela Musteata, Sara De Silvestri, Valerio Bellandi, Paolo Ceravolo, Matteo Cesari, Domenico Azzolino, Alberto Sanna, and Diana Trojaniello. 2022-02-28. Older Adults' and Clinicians' Perspectives on a Smart Health Platform for the Aging Population: Design and Evaluation Study. *JMIR Aging* 5, 1 (2022-02-28), e29623. DOI: http://dx.doi.org/10.2196/29623

[90] Mary Czerwinski, Ran Gilad-Bachrach, Shamsi Iqbal, and Gloria Mark. 2016-09-12. Challenges for designing notifications for affective computing systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. Association for Computing Machinery, 1554–1559. DOI: http://dx.doi.org/10.1145/2968219.2968548

[91] Birgit A. Damoiseaux-Volman, Stephanie Medlock, Marsha D. van der Eijk, Johannes A. Romijn, Ameen Abu-Hanna, and Nathalie van der Velde. 2021-10-01. Falls and delirium in older inpatients: Work-as-imagined, work-as-done and preferences for clinical decision support systems. *Safety Science* 142 (2021-10-01), 105355. DOI: http://dx.doi.org/10.1016/j.ssci.2021.105355

[92] Surabhi Datta, Jordan Godfrey-Stovall, and Kirk Roberts. 2020. RadLex Normalization in Radiology Reports. *AMIA ... Annual Symposium proceedings. AMIA Symposium* 2020 (2020), 338–347.

[93] Jennifer L. Davidson and Carlos Jensen. 2013. Participatory design with older adults: an analysis of creativity in the design of mobile healthcare applications. In *Proceedings of the 9th ACM Conference on Creativity and Cognition (C and C '13)*. Association for Computing Machinery, New York, NY, USA, 114–123. DOI:http://dx.doi.org/10.1145/2466627.2466652

[94] D. W. De Boo, M. Prokop, M. Uffmann, B. van Ginneken, and C. M. Schaefer-Prokop. 2009-11-01. Computer-aided detection (CAD) of lung nodules and small tumours

on chest radiographs. *European Journal of Radiology* 72, 2 (2009-11-01), 218–225. DOI: http://dx.doi.org/10.1016/j.ejrad.2009.05.062

[95] Andrew Dearden, RUdra Joshi, Alexander Figacz, and Sayf Al-Katib. 2022-05-01. Using AutoHotkey Scripts to Perform Voice-Commanded Calculations for Dictated Radiology Reports. *Conference Presentation Abstracts* (2022-05-01). https://scholarlyworks.beaumont.org/radiology_confabstract/78

[96] Sidney Dekker. 2006. *The Field Guide to Understanding Human Error*. Ashgate Publishing, Ltd. Google-Books-ID: NTP5iLn4XHkC.

[97] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012-05-05. "Yours is better!": participant response bias in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1321–1330. DOI: http://dx.doi.org/10.1145/2207676.2208589

[98] Priya Deshpande, Alexander Rasin, Eli Brown, Jacob Furst, Daniela S. Raicu, Steven M. Montner, and Samuel G. Armato. 2018-10. Big Data Integration Case Study for Radiology Data Sources. In *2018 IEEE Life Sciences Conference (LSC)*. 195–198. DOI: http://dx.doi.org/10.1109/LSC.2018.8572185

[99] Audrey Desjardins and Aubree Ball. 2018-06-08. Revealing Tensions in Autobiographical Design in HCI. In *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, 753–764. DOI:http://dx.doi.org/10.1145/3196709.3196781

[100] Anouk T. Dev, Teresa L. Kauf, Amany Zekry, Keyur Patel, Karen Heller, Kevin A. Schulman, and John G. McHutchison. 2008-10-08. Factors influencing the participation of gastroenterologists and hepatologists in clinical research. *BMC health services research* 8 (2008-10-08), 208. DOI:http://dx.doi.org/10.1186/1472-6963-8-208

[101] Sara Donetto, Paola Pierri, Vicki Tsianakas, and Glenn Robert. 2015-06-01. Experience-based Co-design and Healthcare Improvement: Realizing Participatory Design in the Public Sector. *The Design Journal* 18, 2 (2015-06-01), 227–248. DOI:http://dx.doi.org/10.2752/175630615X14212498964312

[102] Clare Donohoe. 2019-01-28. An exploration of NHS clinical staff perceptions of changes to clinical products and their procurement. *Nursing Management* 26, 1 (2019-01-28), 26–33. DOI:http://dx.doi.org/10.7748/nm.2019.e1746

[103] Paul Dourish, Janet Finlay, Phoebe Sengers, and Peter Wright. 2004-04-24. Reflective HCI: towards a critical technical practice. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. Association for Computing Machinery, 1727–1728. DOI:http://dx.doi.org/10.1145/985921.986203

[104] Frank A. Drews and David Fawcett. 2010-09-01. Why Healthcare is Not like Aviation: Control of Natural and Technical Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54, 4 (2010-09-01), 369–373. DOI: http://dx.doi.org/10.1177/154193121005400421 Publisher: SAGE Publications Inc.

[105] Bartłomiej Drop, Katarzyna Drop, Agnieszka Barańska, and Mariola Janiszewska. 2018. Why Is Digitalisation a Need? Chosen Disadvantages of Non-Digital Radiology Systems. *Collegium of Economic Analysis Annals* 52 (2018), 11–26. https://ideas.repec.org//a/sgh/annals/i52y2018p11-26.html Publisher: Warsaw School of Economics, Collegium of Economic Analysis.

[106] Francis T. Durso and Frank A. Drews. 2010-04-01. Health Care, Aviation, and Ecosystems: A Socio-Natural Systems Perspective. *Current Directions in Psychological Science* 19, 2 (2010-04-01), 71–75. DOI:http://dx.doi.org/10.1177/0963721410364728 Publisher: SAGE Publications Inc.

[107] D Eason, S Chakraverty, and J A W Wildsmith. 2011-05-01. Sedation in a radiology department – do radiologists follow their own guidelines? *Scottish Medical Journal* 56, 2 (2011-05-01), 61–63. DOI:http://dx.doi.org/10.1258/smj.2010.010013 Publisher: SAGE Publications.

[108] Emmanuel Kobina Mesi Edzie, Klenam Dzefi-Tettey, Abdul Raman Asemah, Edmund Kwakye Brakohiapa, Samuel Asiamah, Frank Quarshie, Adu Tutu Amankwa, Amrit Raj, Obed Nimo, Evans Boadi, Joshua Mensah Kpobi, Richard Ato Edzie, Bernard Osei, Veronica Turkson, and Henry Kusodzi. 2023-05. Perspectives of radiologists in Ghana about the emerging role of artificial intelligence in radiology. *Heliyon* 9, 5 (2023-05), e15558. DOI:http://dx.doi.org/10.1016/j.heliyon.2023.e15558

[109] Farzad Ehsani, Jim Kimzey, Elaine Zuber, Demitrios Master, and Karen Sudre. 2008-08. Speech to Speech Translation for Nurse Patient Interaction. In *Coling 2008: Proceedings of the workshop on Speech Processing for Safety Critical Translation and Pervasive Applications*, Pierrette Bouillon, Farzad Ehsani, Robert Frederking,

Michael McTear, and Manny Rayner (Eds.). Coling 2008 Organizing Committee, 54–59. https://aclanthology.org/W08-1510

[110] Gunnar Ellingsen and Kristoffer Røed. 2010-12-01. The Role of Integration in Health-Based Information Infrastructures. *Computer Supported Cooperative Work (CSCW)* 19, 6 (2010-12-01), 557–584. DOI:http://dx.doi.org/10.1007/s10606-010-9122-y

[111] Mica R. Endsley and Debra G. Jones. 2001-10-01. Disruptions, Interruptions and Information Attack: Impact on Situation Awareness and Decision Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 45, 2 (2001-10-01), 63–67. DOI:http://dx.doi.org/10.1177/154193120104500214 Publisher: SAGE Publications Inc.

[112] NHS England. 2025. Diagnostic Imaging Dataset Statistics. (2025).

[113] Jean-Francois Esculier, Christian Barton, Rod Whiteley, and Christopher Napier. 2018-12-01. Involving clinicians in sports medicine and physiotherapy research: 'design thinking' to help bridge gaps between practice and evidence. *British Journal of Sports Medicine* 52, 24 (2018-12-01), 1550–1551. DOI:http://dx.doi.org/10.1136/bjsports-2018-100078 Publisher: BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine Section: Editorial.

[114] Emma Falkland. 2023. *Cue utilisation and interruptions in high risk, high consequence environments*. thesis.

[115] Nabi Fatahi, Ferid Krupic, and Mikael Hellström. 2015-06-01. Quality of radiologists' communication with other clinicians—As experienced by radiologists. *Patient Education and Counseling* 98, 6 (2015-06-01), 722–727. DOI:http://dx.doi.org/10.1016/j.pec.2015.02.009

[116] Ross W. Filice and Charles E. Kahn. 2019-04-01. Integrating an Ontology of Radiology Differential Diagnosis with ICD-10-CM, RadLex, and SNOMED CT. *Journal of Digital Imaging* 32, 2 (2019-04-01), 206–210. DOI:http://dx.doi.org/10.1007/s10278-019-00186-3

[117] Ross W. Filice and Charles E. Kahn. 2021-12-01. Biomedical Ontologies to Guide AI Development in Radiology. *Journal of Digital Imaging* 34, 6 (2021-12-01), 1331–1341. DOI:http://dx.doi.org/10.1007/s10278-021-00527-1

[118] Geraldine Fitzpatrick and Gunnar Ellingsen. 2013-08-01. A Review of 25 Years of CSCW Research in Healthcare: Contributions, Challenges and Future Agendas. *Computer Supported Cooperative Work (CSCW)* 22, 4 (2013-08-01), 609–665. DOI:http://dx.doi.org/10.1007/s10606-012-9168-0

[119] Les R. Folio, Laura B. Machado, and Andrew J. Dwyer. 2018-03. Multimedia-enhanced Radiology Reports: Concept, Components, and Challenges. *RadioGraphics* 38, 2 (2018-03), 462–482. DOI:http://dx.doi.org/10.1148/rg.2017170047 Publisher: Radiological Society of North America.

[120] Dominic Furniss. 2014. HCI Observations on an Oncology Ward: A Fieldworker's Experience. In *Fieldwork for Healthcare: Case Studies Investigating Human Factors in Computing Systems*. Springer International Publishing, 19–25. DOI:http://dx.doi.org/10.1007/978-3-031-01596-0_3

[121] Dominic Furniss, Nick Barber, Imogen Lyons, Lina Eliasson, and Ann Blandford. 2014-02. Unintentional non-adherence: can a spoon full of resilience help the medicine go down? *BMJ Quality & Safety* 23, 2 (2014-02), 95–98. DOI:http://dx.doi.org/10.1136/bmjqs-2013-002276

[122] Dominic Furniss, Ann Blandford, and Astrid Mayer. 2011-07-01. Unremarkable errors: low-level disturbances in infusion pump use. *Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction* (2011-07-01).

[123] Dhakshinamoorthy Ganeshan, Phuong-Anh Thi Duong, Linda Probyn, Leon Lenchik, Tatum A. McArthur, Michele Retrouvey, Emily H. Ghobadi, Stephane L. Desouches, David Pastel, and Isaac R. Francis. 2018-01-01. Structured Reporting in Radiology. *Academic Radiology* 25, 1 (2018-01-01), 66–73. DOI:http://dx.doi.org/10.1016/j.acra.2017.08.005

[124] Harold Garfinkel. 1964-01. Studies of the Routine Grounds of Everyday Activities. *Social Problems* 11, 3 (1964-01), 225–250. DOI:http://dx.doi.org/10.2307/798722

[125] Susan A. Gelman and Cristine H. Legare. 2011-10-21. Concepts and Folk Theories. *Annual Review of Anthropology* 40, 1 (2011-10-21), 379–398. DOI:http://dx.doi.org/10.1146/annurev-anthro-081309-145822

[126] Chris Giordano, Meghan Brennan, Basma Mohamed, Parisa Rashidi, François Modave, and Patrick Tighe. 2021. Accessing Artificial Intelligence for Clinical Decision-Making. *Frontiers in Digital Health* 3 (2021). https://www.frontiersin.org/article/10.3389/fdgth.2021.645232

[127] Shiyoh Goetsu and Tetsuya Sakai. 2019-10-14. Voice Input Interface Failures and Frustration: Developer and User Perspectives. In *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, 24–26. DOI: http://dx.doi.org/10.1145/3332167.3357103

[128] Ben Goldacre and Brian MacKenna. 2020-07-17. The NHS deserves better use of hospital medicines data. *BMJ* (2020-07-17), m2607. DOI:http://dx.doi.org/10.1136/bmj.m2607

[129] Patrick Gontar, Sonja Anna Elisabeth Schneider, Carsten Schmidt-Moll, Celina Bollin, and Klaus Bengler. 2017-11-01. Hate to interrupt you, but... analyzing turn-arounds from a cockpit perspective. *Cognition, Technology & Work* 19, 4 (2017-11-01), 837–853. DOI:http://dx.doi.org/10.1007/s10111-017-0440-4

[130] Foster R. Goss, Suzanne V. Blackley, Carlos A. Ortega, Leigh T. Kowalski, Adam B. Landman, Chen-Tan Lin, Marie Meteer, Samantha Bakes, Stephen C. Gradwohl, David W. Bates, and Li Zhou. 2019-10-01. A clinician survey of using speech recognition for clinical documentation in the electronic health record. *International Journal of Medical Informatics* 130 (2019-10-01), 103938. DOI:http://dx.doi.org/10.1016/j.ijmedinf.2019.07.017

[131] UK Government. 2025-02-17. Regulating medical devices in the UK. (2025-02-17). https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk

[132] T. Greenhalgh, S. Hinder, K. Stramer, T. Bratan, and J. Russell. 2010-11-16. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. *BMJ* 341 (2010-11-16), c5814–c5814. Issue nov16 1. DOI:http://dx.doi.org/10.1136/bmj.c5814

[133] Trisha Greenhalgh, Henry W.w. Potts, Geoff Wong, Pippa Bark, and Deborah Swinglehurst. 2009. Tensions and Paradoxes in Electronic Patient Record Research: A Systematic Literature Review Using the Meta-narrative Method. *The Milbank Quarterly*

87, 4 (2009), 729–788. DOI:http://dx.doi.org/10.1111/j.1468-0009.2009.00578.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0009.2009.00578.x.

[134] Tobias Grundgeiger, Jörn Hurtienne, and Oliver Happel. 2021-08-01. Why and How to Approach User Experience in Safety-Critical Domains: The Example of Health Care. *Human Factors* 63, 5 (2021-08-01), 821–832. DOI:http://dx.doi.org/10.1177/0018720819887575 Publisher: SAGE Publications Inc.

[135] Tobias Grundgeiger and Penelope Sanderson. 2009-05-01. Interruptions in healthcare: Theoretical views. *International Journal of Medical Informatics* 78, 5 (2009-05-01), 293–307. DOI:http://dx.doi.org/10.1016/j.ijmedinf.2008.10.001

[136] Normunds Gruzitis, Roberts Dargis, Viesturs Julijs Lasmanis, Ginta Garkaje, and Didzis Gosko. 2022. Adapting Automatic Speech Recognition to the Radiology Domain for a Less-Resourced Language: The Case of Latvian. In *Intelligent Sustainable Systems (Lecture Notes in Networks and Systems)*, Atulya K. Nagar, Dharm Singh Jat, Gabriela Marín-Raventós, and Durgesh Kumar Mishra (Eds.). Springer Nature, 267–276. DOI:http://dx.doi.org/10.1007/978-981-16-6309-3_27

[137] Erik Grönvall and Morten Kyng. 2013-11-01. On participatory design of home-based healthcare. *Cognition, Technology & Work* 15, 4 (2013-11-01), 389–401. DOI: http://dx.doi.org/10.1007/s10111-012-0226-7

[138] Sun Guang-li, Huang Wei, and Li Ji. 2011-10. Improved VUI system based on maintenance device. In *2011 International Conference on Computational Problem-Solving (ICCP)*. IEEE, 510–513. DOI:http://dx.doi.org/10.1109/ICCPS.2011.6092247

[139] B.M. Zeeshan Hameed, Gayathri Prerepa, Vathsala Patil, Pranav Shekhar, Syed Zahid Raza, Hadis Karimi, Rahul Paul, Nithesh Naik, Sachin Modi, Ganesh Vigneswaran, Bhavan Prasad Rai, Piotr Chłosta, and Bhaskar K. Somani. 2021-01-01. Engineering and clinical use of artificial intelligence (AI) with machine learning and data science advancements: radiology leading the way for future. *Therapeutic Advances in Urology* 13 (2021-01-01), 17562872211044880. DOI: http://dx.doi.org/10.1177/17562872211044880 Publisher: SAGE Publications.

[140] Imane Hammana, Luigi Lepanto, Thomas Poder, Christian Bellemare, and My-Sandra Ly. 2015-06-01. Speech Recognition in the Radiology Department: A Systematic

Review. *Health Information Management Journal* 44, 2 (2015-06-01), 4–10. DOI:http://dx.doi.org/10.1177/183335831504400201 Publisher: SAGE Publications Ltd STM.

[141] Nicolai Brodersen Hansen, Christian Dindler, Kim Halskov, Ole Sejer Iversen, Claus Bossen, Ditte Amund Basballe, and Ben Schouten. 2020-01-10. How Participatory Design Works: Mechanisms and Effects. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction (OzCHI '19)*. Association for Computing Machinery, 30–41. DOI:http://dx.doi.org/10.1145/3369457.3369460

[142] Anthony Kwame Harrison. 2022. Spending Time Together Outside of the Clubs: Cross-gender Rapport Building and Interview Recruitment in Music Scene Ethnography. *Sociological Inquiry* 92, 4 (2022), 1329–1351. DOI:http://dx.doi.org/10.1111/soin.12462 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/soin.12462.

[143] Abbas Hassan and Khaled El-Rayes. 2020-06-02. Quantifying the interruption impact of activity delays in non-serial repetitive construction projects. *Construction Management and Economics* 38, 6 (2020-06-02), 515–533. DOI:http://dx.doi.org/10.1080/01446193.2019.1657922 Publisher: Routledge _eprint: https://doi.org/10.1080/01446193.2019.1657922.

[144] D. B. Hayt and S. Alexander. 2001-09. The pros and cons of implementing PACS and speech recognition systems. *Journal of Digital Imaging* 14, 3 (2001-09), 149–157. DOI:http://dx.doi.org/10.1007/s10278-001-0014-z

[145] Elizabeth M. Hecht, John Simmons, and Martin R. Prince. 2015-11-01. Reducing Interruptions in the Reading Room: Standardized CT/MRI Contrast Orders. *Journal of the American College of Radiology* 12, 11 (2015-11-01), 1196–1199. DOI:http://dx.doi.org/10.1016/j.jacr.2015.03.033 Publisher: Elsevier.

[146] Brigida Hernandez, Katherine McDonald, Marielle Divilbiss, Elizabeth Horin, Jessica Velcoff, and Oscar Donoso. 2008-09-01. Reflections from Employers on the Disabled Workforce: Focus Groups with Healthcare, Hospitality and Retail Administrators. *Employee Responsibilities and Rights Journal* 20, 3 (2008-09-01), 157–164. DOI:http://dx.doi.org/10.1007/s10672-008-9063-5

[147] Julia Hirschberg, Diane Litman, and Marc Swerts. 2004-06-01. Prosodic and other cues to speech recognition failures. *Speech Communication* 43, 1 (2004-06-01), 155–175. DOI:http://dx.doi.org/10.1016/j.specom.2004.01.006

[148] Harry Hochheiser and Jonathan Lazar. 2007-12-01. HCI and Societal Issues: A Framework for Engagement. *Int. J. Hum. Comput. Interaction* 23 (2007-12-01), 339–374. DOI:http://dx.doi.org/10.1080/10447310701702717

[149] Anna Hohm, Oliver Happel, Jörn Hurtienne, and Tobias Grundgeiger. 2022-05-01. User experience in safety–critical domains: a survey on motivational orientations and psychological need satisfaction in acute care. *Cognition, Technology & Work* 24, 2 (2022-05-01), 247–260. DOI:http://dx.doi.org/10.1007/s10111-022-00697-0

[150] Erik Hollnagel. 1993. THE PHENOTYPE OF ERRONEOUS ACTIONS: IMPLICATIONS FOR HCI DESIGN. *International Journal of Man-Machine Studies* 39 (1993).

[151] Erik Hollnagel and Robyn Clay-Williams. 2022. Work-as-Imagined and Work-as-Done. In *Implementation Science*. Routledge. Num Pages: 3.

[152] Susan G. Hopkinson and Bonnie Mowinski Jennings. 2013. Interruptions during nurses' work: A state-of-the-science review. *Research in Nursing & Health* 36, 1 (2013), 38–53. DOI:http://dx.doi.org/10.1002/nur.21515 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nur.21515.

[153] Steven Houben, Mads Frost, and Jakob E. Bardram. 2015-02-28. Collaborative Affordances of Hybrid Patient Record Technologies in Medical Work. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. Association for Computing Machinery, 785–797. DOI:http://dx.doi.org/10.1145/2675133.2675164

[154] Hian Liang Huang, Robert Chun Chen, Irene Teo, Isha Chaudhry, Ai Ling Heng, Kun Da Zhuang, Hiang Khoon Tan, and Bien Soo Tan. 2021. A survey of anxiety and burnout in the radiology workforce of a tertiary hospital during the COVID-19 pandemic. *Journal of Medical Imaging and Radiation Oncology* 65, 2 (2021), 139–145. DOI:http://dx.doi.org/10.1111/1754-9485.13152 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1754-9485.13152.

[155] John Hughes, Val King, Tom Rodden, and Hans Andersen. 1994-10-22. Moving out from the control room: ethnography in system design. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94)*. Association for Computing Machinery, 429–439. DOI:http://dx.doi.org/10.1145/192844.193065

[156] Sindre Høyland. 2007-01-01. A theoretical understanding of safe work practices - A comparison of aviation and health care. In *Proceedings of the European Safety and Reliability Conference 2007, ESREL 2007 - Risk, Reliability and Societal Safety*. Vol. 3. 2121–2128. Journal Abbreviation: Proceedings of the European Safety and Reliability Conference 2007, ESREL 2007 - Risk, Reliability and Societal Safety.

[157] Fernando De Castro Guimarães Rios Ignácio, Luis Ronan Marquez Ferreira De Souza, Giuseppe D'Ippolito, and Mayara Martins Garcia. 2018-10-18. Radiology report: what is the opinion of the referring physician? *Radiologia Brasileira* 51, 5 (2018-10-18), 308–312. DOI:http://dx.doi.org/10.1590/0100-3984.2017.0115

[158] Daniel Jarrett, Eleanor Stride, Katherine Vallis, and Mark J. Gooding. 2019-08. Applications and limitations of machine learning in radiation oncology. *The British Journal of Radiology* 92, 1100 (2019-08), 20190001. DOI:http://dx.doi.org/10.1259/bjr.20190001

[159] Seoungdeok Jeon, Zachary Colburn, Joshua Sakai, Ling-Hong Hung, and Ka Yee Yeung. 2021-08-01. Application of natural language processing and machine learning to radiology reports. In *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '21)*. Association for Computing Machinery, 1. DOI:http://dx.doi.org/10.1145/3459930.3469496

[160] Jiepu Jiang, Wei Jeng, and Daqing He. 2013-07-28. How do users respond to voice input errors?: lexical and phonetic query reformulation in voice search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 143–152. DOI:http://dx.doi.org/10.1145/2484028.2484092

[161] Rose Johnson, Kenton O'Hara, Abigail Sellen, Claire Cousins, and Antonio Criminisi. 2011. Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3323–3332. DOI:http://dx.doi.org/10.1145/1978942.1979436

[162] Narinder Kapur, Anam Parand, Tayana Soukup, Tom Reader, and Nick Sevdalis. 2016-01-01. Aviation and healthcare: a comparative review with implications for patient safety. *JRSM Open* 7, 1 (2016-01-01), 2054270415616548. DOI:http://dx.doi.org/10.1177/2054270415616548 Publisher: SAGE Publications.

[163] Nicholas Kasdaglis and Kimberly Stowers. 2016. Beyond Human Factors: The Role of Human Centered Design in Developing a Safety-Critical System. In *HCI International 2016 – Posters' Extended Abstracts (Communications in Computer and Information Science)*, Constantine Stephanidis (Ed.). Springer International Publishing, 345–351. DOI: http://dx.doi.org/10.1007/978-3-319-40548-3_58

[164] Sarah Kettley, Richard Kettley, and Matthew Bates. 2015-09-07. An introduction to the person-centred approach as an attitude for participatory design. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct)*. Association for Computing Machinery, 1101–1104. DOI:http://dx.doi.org/10.1145/2800835.2807945

[165] Hanneke Kip, Julia Keizer, Marcia C. da Silva, Nienke Beerlage-de Jong, Nadine Köhle, and Saskia M. Kelders. 2022-01-27. Methods for Human-Centered eHealth Development: Narrative Scoping Review. *Journal of Medical Internet Research* 24, 1 (2022-01-27), e31858. DOI:http://dx.doi.org/10.2196/31858 Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

[166] Susanne Kirchner, Jessica Schroeder, James Fogarty, and Sean A. Munson. 2021-05-06. "They don't always think about that": Translational Needs in the Design of Personal Health Informatics Applications. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, 1–16. DOI:http://dx.doi.org/10.1145/3411764.3445587

[167] John C. Knight. 2002-05-19. Safety critical systems: challenges and directions. In *Proceedings of the 24th International Conference on Software Engineering (ICSE '02)*. Association for Computing Machinery, 547–550. DOI:http://dx.doi.org/10.1145/581339.581406

[168] Leo Kobayashi, I. I. I. Robert M. Boss, Frantz J. Gibbs, Eric Goldlust, Michelle M. Hennedy, James E. Monti, and Nathan A. Siegel. 2011-07-01. Color-Coding and Human Factors Engineering to Improve Patient Safety Characteristics of Paper-Based Emergency Department Clinical Documentation. *HERD: Health Environments Research*

*& Design Journal* (2011-07-01). DOI:http://dx.doi.org/10.1177/193758671100400406 Publisher: SAGE PublicationsSage CA: Los Angeles, CA.

[169] Marc Kohli, Luciano M. Prevedello, Ross W. Filice, and J. Raymond Geis. 2017-04. Implementing Machine Learning in Radiology Practice and Research. *American Journal of Roentgenology* 208, 4 (2017-04), 754–760. DOI:http://dx.doi.org/10.2214/AJR.16.17224

[170] Alexander Komashie, Guillaume Lamé, François Patou, Nicholas Ciccone, Anja Maier, and P. Clarkson. 2019-07-01. Exploring Healthcare Systems Design Research and Practice: Outcomes of an International Meeting, Vol. 1. 947–956. DOI:http://dx.doi.org/10.1017/dsi.2019.100

[171] Sara A. Kreindler, Ashley Struthers, Colleen J. Metge, Catherine Charette, Karen Harlos, Paul Beaudin, Sunita B. Bapuji, Ingrid Botting, and Jose Francois. 2019-03-28. Pushing for partnership: physician engagement and resistance in primary care renewal. *Journal of Health Organization and Management* 33, 2 (2019-03-28), 126–140. DOI:http://dx.doi.org/10.1108/JHOM-05-2018-0141

[172] Jennifer Lai and John Vergo. 1997-03-27. MedSpeak: report creation with continuous speech recognition. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. ACM, 431–438. DOI:http://dx.doi.org/10.1145/258549.258829

[173] H-L Langen, J Bielmeier, G Wittenberg, R Selbach, and H Feustel. 2003-10-01. [Workflow improvement and efficiency gain with near total digitalization of a radiology department]. *RoFo* 175, 10 (2003-10-01), 1309–1316. DOI:http://dx.doi.org/10.1055/s-2003-42889

[174] David B. Larson, Craig M. Froehle, Neil D. Johnson, and Alexander J. Towbin. 2014-11. Communication in Diagnostic Radiology: Meeting the Challenges of Complexity. *American Journal of Roentgenology* 203, 5 (2014-11), 957–964. DOI:http://dx.doi.org/10.2214/AJR.14.12949

[175] W. Larsson, P. Aspelin, M. Bergquist, K. Hillergård, B. Jacobsson, L. Lindsköld, J. Wallberg, and N. Lundberg. 2007-08-01. The effects of PACS on radiographer's work practice. *Radiography* 13, 3 (2007-08-01), 235–240. DOI:http://dx.doi.org/10.1016/j.radi.2006.02.005

[176] Byung Cheol Lee and Vincent G. Duffy. 2015. The Effects of Task Interruption on Human Performance: A Study of the Systematic Classification of Human Behavior and Interruption Frequency. *Human Factors and Ergonomics in Manufacturing & Service Industries* 25, 2 (2015), 137–152. DOI:http://dx.doi.org/10.1002/hfm.20603

[177] John D. Lee. 2014-03. Dynamics of Driver Distraction: The process of engaging and disengaging. *Annals of Advances in Automotive Medicine* 58 (2014-03), 24–32. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001670/

[178] Yura Lee, Min-Young Jung, Gee Won Shin, Sangwoo Bahn, Taezoon Park, Insook Cho, and Jae-Ho Lee. 2018-07-31. Safety and Usability Guidelines of Clinical Information Systems Integrating Clinical Workflow: A Systematic Review. *Healthcare Informatics Research* 24, 3 (2018-07-31), 157–169. DOI:http://dx.doi.org/10.4258/hir.2018.24.3.157 Publisher: Korean Society of Medical Informatics.

[179] Nancy Leveson and Clark Turner. 1992. *An Investigation of the THERAC-25 Accidents*. University of California, Irvine.

[180] Yunqiu Li, Patrick Oladimeji, Carlos Monroy, Abigail Cauchi, Harold Thimbleby, Dominic Furniss, Chris Vincent, and Ann Blandford. 2011-12. Design of interactive medical devices: Feedback and its improvement. In *2011 IEEE International Symposium on IT in Medicine and Education*, Vol. 2. 204–208. DOI:http://dx.doi.org/10.1109/ITiME.2011.6132022

[181] Calvin A. Liang, Sean A. Munson, and Julie A. Kientz. 2021-04-30. Embracing Four Tensions in Human-Computer Interaction Research with Marginalized People. *ACM Transactions on Computer-Human Interaction* 28, 2 (2021-04-30), 1–47. DOI: http://dx.doi.org/10.1145/3443686

[182] Joseph Lindley. 2015-06-22. Researching Design Fiction With Design Fiction. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. ACM, 325–326. DOI:http://dx.doi.org/10.1145/2757226.2764763

[183] Oskar Lindwall and Gustav Lymer. 2005. Vulgar competence, ethnomethodological indifference and curricular design. 388–397. DOI:http://dx.doi.org/10.3115/1149293.1149344

[184] P. Littlejohns. 2003-04-19. Evaluating computerised health information systems: hard lessons still to be learnt. *BMJ* 326, 7394 (2003-04-19), 860–863. DOI:http://dx.doi.org/10.1136/bmj.326.7394.860

[185] Yikun Liu, Yuan Jia, Wei Pan, and Mark S. Pfaff. 2014-02-15. Supporting task resumption using visual feedback. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*. Association for Computing Machinery, 767–777. DOI:http://dx.doi.org/10.1145/2531602.2531710

[186] Duri Long and Brian Magerko. 2020-04-21. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–16. DOI:http://dx.doi.org/10.1145/3313831.3376727

[187] F. L. Lucas. 2012. *Style: The Art of Writing Well* (3rd ed ed.). Harriman House.

[188] Paul Luff, Jon Hindmarsh, and Christian Heath. 2000. Workplace studies: recovering work practice and informing system design. (2000).

[189] Paul Luff, Jon Hindmarsh, Christian Heath, and Pamela Hinds. 2003-01-01. Workplace Studies: Recovering Work Practice and Informing System Design. *Computer Supported Cooperative Work* 12 (2003-01-01), 123–125. DOI:http://dx.doi.org/10.1017/CBO9780511628122

[190] Melissa Luke and Kristopher M. Goodrich. 2019-04-03. Focus Group Research: An Intentional Strategy for Applied Group Research? *The Journal for Specialists in Group Work* 44, 2 (2019-04-03), 77–81. DOI:http://dx.doi.org/10.1080/01933922.2019.1603741

[191] Jessica Lynn Campbell. 2020-10-03. A Mixed-methods Approach to Evaluating the Usability of Telemedicine Communications. In *Proceedings of the 38th ACM International Conference on Design of Communication*. ACM, 1–6. DOI:http://dx.doi.org/10.1145/3380851.3416755

[192] Ritch Macefield. 2007. Usability studies and the Hawthorne effect. *J. Usability Studies* 2, 3 (May 2007), 145–154.

[193] Bronislaw Malinowski. 2014. *Argonauts of the Western Pacific: an account of Native Enterprise and adventure in the Archipelagos of Melanesian New Guinea*. Routledge.

[194] Steven C. Mallam and Monica Lundh. 2013-09-01. Ship Engine Control Room Design: Analysis of Current Human Factors & Ergonomics Regulations & Future Directions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1 (2013-09-01), 521–525. DOI:http://dx.doi.org/10.1177/1541931213571112 Publisher: SAGE Publications Inc.

[195] Kirsti Malterud and Kari Tove Elvbakken. 2019-07-18. Patients participating as co-researchers in health research: A systematic review of outcomes and experiences. *Scandinavian Journal of Public Health* (2019-07-18). DOI:http://dx.doi.org/10.1177/1403494819863514 Publisher: SAGE PublicationsSage UK: London, England.

[196] Philippe Marrast, Pascale Zaraté, and Anne Mayère. 2013-02-23. How to support coordination through annotations? A longitudinal case study of nurses' work in an oncology hospital. 207–212. DOI:http://dx.doi.org/10.1145/2441955.2442007

[197] Elizabeth Mattox. 2012-08-01. Medical Devices and Patient Safety. *Critical care nurse* 32 (2012-08-01), 60–8. DOI:http://dx.doi.org/10.4037/ccn2012925

[198] Kirsty McNeil, Carina Banziger, and Ian Zealley. 2022-09-01. Reducing interruptions during radiology reporting: Beneficial effect of duty radiologist and office reporting. *Clinical Radiology* 77 (2022-09-01), e26. DOI:http://dx.doi.org/10.1016/j.crad.2022.08.106 Publisher: Elsevier.

[199] Lisa K. Meloncon. 2017-08-04. Patient experience design: expanding usability methodologies for healthcare. *Communication Design Quarterly* 5, 2 (2017-08-04), 19–28. DOI:http://dx.doi.org/10.1145/3131201.3131203

[200] Tilo Mentler, Philippe Palanque, Susanne Boll, Chris Johnson, and Kristof Van Laerhoven. 2021. Control Rooms in Safety Critical Contexts: Design, Engineering and Evaluation Issues. In *Human-Computer Interaction – INTERACT 2021*. 530–535. DOI:http://dx.doi.org/10.1007/978-3-030-85607-6_72

[201] Tilo Mentler, Philippe Palanque, Kristof Van Laerhoven, Margareta Holtensdotter Lützhöft, and Nadine Flegel. 2023. On Land, at Sea, and in the Air: Human-Computer Interaction in Safety-Critical Spaces of Control. In *Human-Computer Interaction – INTERACT 2023*, José Abdelnour Nocera, Marta Kristín Lárusdóttir, Helen Petrie, Antonio Piccinno, and Marco Winckler (Eds.). Springer Nature Switzerland, 657–661. DOI:http://dx.doi.org/10.1007/978-3-031-42293-5_89

[202] Laura G Militello, Nicole B Arbuckle, Jason J Saleem, Emily Patterson, Mindy Flanagan, David Haggstrom, and Bradley N Doebbeling. 2014-03-01. Sources of variation in primary care clinical workflow: Implications for the design of cognitive support. *Health Informatics Journal* 20, 1 (2014-03-01), 35–49. DOI:http://dx.doi.org/10.1177/1460458213476968 Publisher: SAGE Publications Ltd.

[203] R. Minhas. 2007. Does copying clinical or sharing correspondence to patients result in better care? *International Journal of Clinical Practice* 61, 8 (2007), 1390–1395. DOI: http://dx.doi.org/10.1111/j.1742-1241.2007.01432.x

[204] Torin Monahan and Jill A. Fisher. 2010-06. Benefits of 'observer effects': lessons from the field. *Qualitative Research* 10, 3 (2010-06), 357–376. DOI:http://dx.doi.org/10.1177/1468794110362874

[205] Christopher A. Monk, Deborah A. Boehm-Davis, George Mason, and J. Gregory Trafton. 2004-12-01. Recovering From Interruptions: Implications for Driver Distraction Research. *Human Factors* 46, 4 (2004-12-01), 650–663. DOI:http://dx.doi.org/10.1518/hfes.46.4.650.56816 Publisher: SAGE Publications Inc.

[206] Emmanuel Montagnon, Milena Cerny, Alexandre Cadrin-Chênevert, Vincent Hamilton, Thomas Derennes, André Ilinca, Franck Vandenbroucke-Menu, Simon Turcotte, Samuel Kadoury, and An Tang. 2020-02-10. Deep learning workflow in radiology: a primer. *Insights into Imaging* 11, 1 (2020-02-10), 22. DOI: http://dx.doi.org/10.1186/s13244-019-0832-5

[207] Rebecca L. Morris, Angela Ruddock, Kay Gallacher, Carly Rolfe, Sally Giles, and Stephen Campbell. 2021. Developing a patient safety guide for primary care: A co-design approach involving patients, carers and clinicians. *Health Expectations* 24, 1 (2021), 42–52. DOI:http://dx.doi.org/10.1111/hex.13143 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hex.13143.

[208] R. E. Motyer, S. Liddy, W. C. Torreggiani, and O. Buckley. 2016-11-01. Frequency and analysis of non-clinical errors made in radiology reports using the National Integrated Medical Imaging System voice recognition dictation software. *Irish Journal of Medical Science (1971 -)* 185, 4 (2016-11-01), 921–927. DOI:http://dx.doi.org/10.1007/s11845-016-1507-6

[209] Siddhartha Mukherjee. 2017. A.I. Versus M.D. *The New Yorker* (2017).

[210] Michael Muller, Jeffrey Bardzell, EunJeong Cheon, Norman Makoto Su, Eric P. S. Baumer, Casey Fiesler, Ann Light, and Mark Blythe. 2020-04-25. Understanding the Past, Present, and Future of Design Fictions. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–8. DOI:http://dx.doi.org/10.1145/3334480.3375168

[211] Reese Muntean, Kate Hennessy, Chelsea Mills, and Alissa N. Antle. 2024-07-01. Untangling Cables: A Case Study of the Life & Afterlife of Digital Devices in Academic Research. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, 653–667. DOI:http://dx.doi.org/10.1145/3643834.3661540

[212] Daniel R. Murphy, Hardeep Singh, and Leonard Berlin. 2014-12-01. Communication breakdowns and diagnostic errors: a radiology perspective. *Diagnosis* 1, 4 (2014-12-01), 253–261. DOI:http://dx.doi.org/10.1515/dx-2014-0035

[213] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018-04-19. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, 1–7. DOI:http://dx.doi.org/10.1145/3173574.3173580

[214] Chelsea M. Myers, Anushay Furqan, and Jichen Zhu. 2019-05-02. The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, 1–9. DOI:http://dx.doi.org/10.1145/3290605.3300277

[215] Chelsea M. Myers, Luis Fernando Laris Pardo, Ana Acosta-Ruiz, Alessandro Canossa, and Jichen Zhu. 2021-07-27. "Try, Try, Try Again:" Sequence Analysis of User Interaction Data with a Voice User Interface. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, 1–8. DOI:http://dx.doi.org/10.1145/3469595.3469613

[216] Louise Møller, Karen Winther Christiansen, and Anja Sigaard Mortensen. 2020-03-01. Interruptions Affect the Quality of Radiographic Practice. *Journal of Radiology Nursing* 39, 1 (2020-03-01), 48–52. DOI:http://dx.doi.org/10.1016/j.jradnu.2019.10.003

[217] Francis Müller. 2021. *Design Ethnography: Epistemology and Methodology*. Springer International Publishing. DOI:http://dx.doi.org/10.1007/978-3-030-60396-0

[218] Frederik Naujoks, Katharina Wiedemann, and Nadja Schömig. 2017-09-24. The Importance of Interruption Management for Usefulness and Acceptance of Automated Driving. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17)*. Association for Computing Machinery, 254–263. DOI:http://dx.doi.org/10.1145/3122986.3123000

[219] Emanuele Neri, Francesca Coppola, Vittorio Miele, Corrado Bibbolino, and Roberto Grassi. 2020-06-01. Artificial intelligence: Who is responsible for the diagnosis? *La radiologia medica* 125, 6 (2020-06-01), 517–521. DOI:http://dx.doi.org/10.1007/s11547-020-01135-9

[220] Daniel Neyland and Andrea Whittle. 2018-06-01. Garfinkel on strategy: Using ethnomethodology to make sense of "rubbish strategy". *Critical Perspectives on Accounting* 53 (2018-06-01), 31–42. DOI:http://dx.doi.org/10.1016/j.cpa.2017.03.008

[221] Ming-Yen Ng and Nitin Ramamurthy. 2012. A career in radiology. *BMJ* 345 (2012). DOI:http://dx.doi.org/10.1136/bmj.e8142 Publisher: BMJ Publishing Group Ltd _eprint: https://www.bmj.com/content.

[222] Z. Niazkhani, H. Pirnejad, M. Berg, and J. Aarts. 2009-07-01. The Impact of Computerized Provider Order Entry Systems on Inpatient Clinical Workflow: A Literature Review. *Journal of the American Medical Informatics Association* 16, 4 (2009-07-01), 539–549. DOI:http://dx.doi.org/10.1197/jamia.M2419

[223] Jodi Heintz Obradovich and David D. Woods. 1996-12-01. SPECIAL SECTION: Users as Designers: How People Cope with Poor HCI Design in Computer-Based Medical Devices. *Human Factors* 38, 4 (1996-12-01), 574–592. DOI:http://dx.doi.org/10.1518/001872096778827251 Publisher: SAGE Publications Inc.

[224] Željko Obrenović. 2014-10-30. The Hawthorne studies and their relevance to HCI research. *Interactions* 21, 6 (2014-10-30), 46–51. DOI:http://dx.doi.org/10.1145/2674966

[225] The Royal College of Radiologists. 2023. State of the wait | The Royal College of Radiologists. (2023). https://www.rcr.ac.uk/news-policy/policy-reports-initiatives/state-of-the-wait/

[226] The Royal College of Radiologists. 2024. What does a clinical radiologist do? | The Royal College of Radiologists. (2024). https://www.rcr.ac.uk/our-specialties/clinical-radiology/discover-clinical-radiology/thinking-about-a-career-in-clinical-radiology/what-does-a-clinical-radiologist-do/

[227] University of Stirling. Psychological Image Collection at Stirling (PICS). (????). https://pics.stir.ac.uk/

[228] S. Oviatt and R. VanGent. 1996-10. Error resolution during multimodal human-computer interaction. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, Vol. 1. 204–207 vol.1. DOI:http://dx.doi.org/10.1109/ICSLP.1996.607077

[229] Siobhan B. O' Neill, Arvind Vijayasarathi, Savvas Nicolaou, Frances Walstra, Noriko Salamon, Peter L. Munk, and Faisal Khosa. 2021-11-01. Evaluating Radiology Result Communication in the Emergency Department. *Canadian Association of Radiologists Journal* 72, 4 (2021-11-01), 846–853. DOI:http://dx.doi.org/10.1177/0846537119899268 Publisher: SAGE Publications Inc.

[230] Emre Pakdemirli. 2019-02. Artificial intelligence in radiology: friend or foe? Where are we now and where are we heading? *Acta Radiologica Open* 8, 2 (2019-02), 2058460119830222. DOI:http://dx.doi.org/10.1177/2058460119830222

[231] Elias Parker. 2012-04-24. Creativity-based Research: The Process of Co-Designing with Users. (2012-04-24). https://uxmag.com/articles/creativity-based-research-the-process-of-co-designing-with-users

[232] John Pepper. 2019-07-07. Faulty devices: how does cardiology compare to aviation? *European Heart Journal* 40, 26 (2019-07-07), 2103–2106. DOI:http://dx.doi.org/10.1093/eurheartj/ehz422

[233] Haroldas Petkus, Jan Hoogewerf, and Jeremy C Wyatt. 2020-05. What do senior physicians think about AI and clinical decision support systems: Quantitative and

qualitative analysis of data from specialty societies. *Clinical Medicine* 20, 3 (2020-05), 324–328. DOI:http://dx.doi.org/10.7861/clinmed.2019-0317

[234] Anna Pettican, Beverley Goodman, Wendy Bryant, Peter Beresford, Paul Freeman, Valerie Gladwell, Cherry Kilbride, and Ewen Speed. 2022-11-28. Doing together: reflections on facilitating the co-production of participatory action research with marginalised populations. *Qualitative Research in Sport, Exercise and Health* 15 (2022-11-28), 1–18. DOI:http://dx.doi.org/10.1080/2159676X.2022.2146164

[235] Chiara Pomare, Kate Churruca, Janet C. Long, Louise A. Ellis, and Jeffrey Braithwaite. 2021a. Work-as-Imagined Versus Work-as-Done: The Disconnect Between Policy Expectations and Staff Experiences in Hospital Redevelopment. In *Managing Healthcare Organisations in Challenging Policy Contexts*, Roman Kislov, Diane Burns, Bjørn Erik Mørk, and Kathleen Montgomery (Eds.). Springer International Publishing, 133–149. DOI:http://dx.doi.org/10.1007/978-3-030-81093-1_7

[236] Chiara Pomare, Kate Churruca, Janet C. Long, Louise A. Ellis, and Jeffrey Braithwaite. 2021b. Work-as-Imagined Versus Work-as-Done: The Disconnect Between Policy Expectations and Staff Experiences in Hospital Redevelopment. In *Managing Healthcare Organisations in Challenging Policy Contexts*, Roman Kislov, Diane Burns, Bjørn Erik Mørk, and Kathleen Montgomery (Eds.). Springer International Publishing, 133–149. DOI:http://dx.doi.org/10.1007/978-3-030-81093-1_7 Series Title: Organizational Behaviour in Healthcare.

[237] Ewoud Pons, Loes M. M. Braun, M. G. Myriam Hunink, and Jan A. Kors. 2016-05. Natural Language Processing in Radiology: A Systematic Review. *Radiology* 279, 2 (2016-05), 329–343. DOI:http://dx.doi.org/10.1148/radiol.16142770

[238] Maxine Power, Matthew Fogarty, John Madsen, Katherine Fenton, Kevin Stewart, Ailsa Brotherton, Katherine Cheema, Abigail Harrison, and Lloyd Provost. 2014-06. Learning from the design and development of the NHS Safety Thermometer. *International Journal for Quality in Health Care* 26, 3 (2014-06), 287–297. DOI:http://dx.doi.org/10.1093/intqhc/mzu043

[239] Anna Prenestini, Rocco Palumbo, Roberto Grilli, and Federico Lega. 2023-09-26. Exploring physician engagement in health care organizations: a scoping review.

*BMC Health Services Research* 23, 1 (2023-09-26), 1029. DOI:http://dx.doi.org/10.1186/s12913-023-09935-1

[240] The Royal College Of Radiologists. Clinical radiology census reports | The Royal College of Radiologists. (????). https://www.rcr.ac.uk/news-policy/policy-reports-initiatives/clinical-radiology-census-reports/

[241] The Royal College Of Radiologists. 2024. What does a clinical radiologist do? | The Royal College of Radiologists. (2024). https://www.rcr.ac.uk/our-specialties/clinical-radiology/discover-clinical-radiology/thinking-about-a-career-in-clinical-radiology/what-does-a-clinical-radiologist-do/

[242] Sayeeda Rahman, Md Anwarul Azim Majumder, Sami F Shaban, Nuzhat Rahman, Moslehuddin Ahmed, Khalid Bin Abdulrahman, and Urban JA D'Souza. 2011-03-07. Physician participation in clinical research and trials: issues and approaches. *Advances in Medical Education and Practice* 2 (2011-03-07), 85–93. DOI:http://dx.doi.org/10.2147/AMEP.S14103

[243] Bhargav Raman, Raghav Raman, Lalithakala Raman, and Christopher F. Beaulieu. 2004-01. Radiology on Handheld Devices: Image Display, Manipulation, and PACS Integration Issues. *RadioGraphics* 24, 1 (2004-01), 299–310. DOI:http://dx.doi.org/10.1148/rg.241035127 Publisher: Radiological Society of North America.

[244] D. S. Rana, G. Hurst, L. Shepstone, J. Pilling, J. Cockburn, and M. Crawford. 2005-11-01. Voice recognition for radiology reporting: Is it good enough? *Clinical Radiology* 60, 11 (2005-11-01), 1205–1212. DOI:http://dx.doi.org/10.1016/j.crad.2005.07.002

[245] David Randall, Mark Rouncefield, and Peter Tolmie. 2021-04. Ethnography, CSCW and Ethnomethodology. *Computer Supported Cooperative Work (CSCW)* 30, 2 (2021-04), 189–214. DOI:http://dx.doi.org/10.1007/s10606-020-09388-8

[246] Amon Rapp. 2020-02-01. Design fictions for learning: A method for supporting students in reflecting on technology in Human-Computer Interaction courses. *Computers & Education* 145 (2020-02-01), 103725. DOI:http://dx.doi.org/10.1016/j.compedu.2019.103725

[247] Angela Rayner. 2023. Improving systems: understanding 'work as imagined' versus 'work as done'. *In Practice* 45, 1 (2023), 55–57. DOI:http://dx.doi.org/10.1002/inpr.282 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/inpr.282.

[248] Tom W. Reader and Brian H. Cuthbertson. 2011-11-30. Teamwork and team training in the ICU: Where do the similarities with aviation end? *Critical Care* 15, 6 (2011-11-30), 313. DOI:http://dx.doi.org/10.1186/cc10353

[249] Michael Recht and R. Nick Bryan. 2017-11. Artificial Intelligence: Threat or Boon to Radiologists? *Journal of the American College of Radiology* 14, 11 (2017-11), 1476–1480. DOI:http://dx.doi.org/10.1016/j.jacr.2017.07.007

[250] Madhu Reddy and Paul Dourish. 2002-11-16. A finger on the pulse: temporal rhythms and information seeking in medical work. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work (CSCW '02)*. Association for Computing Machinery, 344–353. DOI:http://dx.doi.org/10.1145/587078.587126

[251] Madhu C. Reddy, Paul Dourish, and Wanda Pratt. 2006-02-01. Temporality in Medical Work: Time also Matters. *Computer Supported Cooperative Work (CSCW)* 15, 1 (2006-02-01), 29–53. DOI:http://dx.doi.org/10.1007/s10606-005-9010-z

[252] Stuart Reeves, Martin Porcheron, and Joel Fischer. 2018-12-21. 'This is not what we wanted': designing for conversation with voice interfaces. *Interactions* 26, 1 (2018-12-21), 46–51. DOI:http://dx.doi.org/10.1145/3296699

[253] John Reicher, Stuart Currie, and Daniel Birchall. 2018-04-01. Safety of working patterns among UK neuroradiologists: what can we learn from the aviation industry and cognitive science? *British Journal of Radiology* 91, 1084 (2018-04-01), 20170284. DOI:http://dx.doi.org/10.1259/bjr.20170284

[254] A. J. Rivera-Rodriguez and B.-T. Karsh. 2010-08-01. Interruptions and distractions in healthcare: review and reappraisal. *Quality and Safety in Health Care* 19, 4 (2010-08-01), 304–312. DOI:http://dx.doi.org/10.1136/qshc.2009.033282

[255] Wendy Roldan, Ziyue Li, Xin Gao, Sarah Kay Strickler, Allison Marie Hishikawa, Jon E. Froehlich, and Jason Yip. 2021-06-28. Pedagogical Strategies for Reflection in Project-based HCI Education with End Users. In *Proceedings of the 2021 ACM Designing*

*Interactive Systems Conference (DIS '21)*. Association for Computing Machinery, 1846–1860. DOI:http://dx.doi.org/10.1145/3461778.3462113

[256] Stephanie Rosenbaum, Gilbert Cockton, Kara Coyne, Michael Muller, and Thyra Rauch. 2002. Focus groups in HCI: wealth of information or waste of resources?. In *CHI '02 extended abstracts on Human factors in computing systems - CHI '02*. ACM Press, 702. DOI:http://dx.doi.org/10.1145/506443.506554

[257] Nithya Sambasivan and Rajesh Veeraraghavan. 2022-04-29. The Deskilling of Domain Expertise in AI Development. In *CHI Conference on Human Factors in Computing Systems*. ACM, 1–14. DOI:http://dx.doi.org/10.1145/3491102.3517578

[258] David Sarkany, Carolynn M. DeBenedectis, and Stephen D. Brown. 2018-07-01. A Review of Resources and Methodologies Available for Teaching and Assessing Patient-Related Communication Skills in Radiology. *Academic Radiology* 25, 7 (2018-07-01), 955–961. DOI:http://dx.doi.org/10.1016/j.acra.2017.11.018

[259] Christina Schneegass and Fiona Draxler. 2021. Designing Task Resumption Cues for Interruptions in Mobile Learning Scenarios. In *Technology-Augmented Perception and Cognition*, Tilman Dingler and Evangelos Niforatos (Eds.). Springer International Publishing, 125–181. DOI:http://dx.doi.org/10.1007/978-3-030-30457-7_5

[260] Christina Schneegass, Vincent Füseschi, Viktoriia Konevych, and Fiona Draxler. 2022-01. Investigating the Use of Task Resumption Cues to Support Learning in Interruption-Prone Environments. *Multimodal Technologies and Interaction* 6, 1 (2022-01), 2. DOI:http://dx.doi.org/10.3390/mti6010002 Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[261] Charlott Sellberg and Jessica Lindblom. 2014-11-01. Comparing methods for workplace studies: a theoretical and empirical analysis. *Cognition, Technology & Work* 16, 4 (2014-11-01), 467–486. DOI:http://dx.doi.org/10.1007/s10111-014-0273-3

[262] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005-08-20. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility (CC '05)*. Association for Computing Machinery, 49–58. DOI:http://dx.doi.org/10.1145/1094562.1094569

[263] Phoebe Sengers, John McCarthy, and Paul Dourish. 2006-04-21. Reflective HCI: articulating an agenda for critical practice. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. Association for Computing Machinery, 1683–1686. DOI:http://dx.doi.org/10.1145/1125451.1125762

[264] National Health Service. 2025. Integrated Research Application System. (2025). https://www.myresearchproject.org.uk/

[265] The Health Informatics Service (THIS). 2021-11-01. How voice recognition technology is benefitting NHS staff and patients. (2021-11-01). http://www.this.nhs.uk/insights/article/voice-recognition-technology

[266] Summit H. Shah, Lamya A. Atweh, Corey A. Thompson, Stephanie Carzoo, Rajesh Krishnamurthy, and Nicholas A. Zumberge. 2022-11-01. Workflow Interruptions and Effect on Study Interpretation Efficiency. *Current Problems in Diagnostic Radiology* 51, 6 (2022-11-01), 848–851. DOI:http://dx.doi.org/10.1067/j.cpradiol.2022.06.003

[267] Shuoqi Shen, Krissy A.R. Doyle-Thomas, Lori Beesley, Amir Karmali, Laura Williams, Nadia Tanel, and Amy C. McPherson. 2017. How and why should we engage parents as co-researchers in health research? A scoping review of current practices. *Health Expectations* 20, 4 (2017), 543–554. DOI:http://dx.doi.org/10.1111/hex.12490 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hex.12490.

[268] Bettina Siewert, Olga R. Brook, Mary Hochman, and Ronald L. Eisenberg. 2016-03. Impact of Communication Errors in Radiology on Patient Care, Customer Satisfaction, and Work-Flow Efficiency. *American Journal of Roentgenology* 206, 3 (2016-03), 573–579. DOI:http://dx.doi.org/10.2214/AJR.15.15117

[269] Bettina Siewert, Suzanne Swedeen, Olga R. Brook, Ronald L. Eisenberg, Lauge Sokol-Hessner, and Jonathan B. Kruskal. 2022-03. Emotional Harm in the Radiology Department: Analysis of an Underrecognized Preventable Error. *Radiology* 302, 3 (2022-03), 613–619. DOI:http://dx.doi.org/10.1148/radiol.2021211846 Publisher: Radiological Society of North America.

[270] Jesper Simonsen and Morten Hertzum. 2012. Sustained Participatory Design: Extending the Iterative Approach. *Design Issues* 28, 3 (2012), 10–21. https://www.jstor.org/stable/23273835 Publisher: The MIT Press.

[271] Aneesha Singh, Nikki Newhouse, Jo Gibbs, Ann E. Blandford, Yunan Chen, Pam Briggs, Helena Mentis, Kate M. Sellen, and Jakob E. Bardram. 2017-05-06. HCI and Health: Learning from Interdisciplinary Interactions. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1322–1325. DOI:http://dx.doi.org/10.1145/3027063.3049287

[272] Hardeep Singh, Ashley N. D. Meyer, and Eric J. Thomas. 2014-09. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ quality & safety* 23, 9 (2014-09), 727–731. DOI:http://dx.doi.org/10.1136/bmjqs-2013-002627

[273] Hardeep Singh, Aanand Dinkar Naik, Raghuram Rao, and Laura Ann Petersen. 2008-04. Reducing diagnostic errors through effective communication: harnessing the power of information technology. *Journal of General Internal Medicine* 23, 4 (2008-04), 489–494. DOI:http://dx.doi.org/10.1007/s11606-007-0393-z

[274] Olli Sjöblom, Juho Heimonen, Lotta Kauhanen, Veronika Laippala, Heljä Lundgrén-Laine, Laura-Maria Murtola, Tapio Salakoski, and Sanna Salanterä. 2012. Avoiding Hazards – What Can Health Care Learn from Aviation?. In *Exploring the Abyss of Inequalities*, Kristina Eriksson-Backa, Annika Luoma, and Erica Krook (Eds.). Springer, 119–127. DOI:http://dx.doi.org/10.1007/978-3-642-32850-3_11

[275] Karin Slegers, Andrea Wilkinson, and Niels Hendriks. 2013-04-27. Active collaboration in healthcare design: participatory design to develop a dementia care app. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. Association for Computing Machinery, 475–480. DOI:http://dx.doi.org/10.1145/2468356.2468440

[276] Vera Sorin, Yiftach Barash, Eli Konen, and Eyal Klang. 2020-05. Deep Learning for Natural Language Processing in Radiology—Fundamentals and a Systematic Review. *Journal of the American College of Radiology* 17, 5 (2020-05), 639–648. DOI: http://dx.doi.org/10.1016/j.jacr.2019.12.026

[277] James P. Spradley. 2016-02-17a. *The Ethnographic Interview*. Waveland Press. Google-Books-ID: KZ3lCwAAQBAJ.

[278] James P. Spradley. 2016-02-17b. *Participant Observation*. Waveland Press. Google-Books-ID: q7DlCwAAQBAJ.

[279] Kirk St. Amant. 2016-01-08. Introduction to the special issue: Cultural considerations for communication design: integrating ideas of culture, communication, and context into user experience design. *Communication Design Quarterly* 4, 1 (2016-01-08), 6–22. DOI:http://dx.doi.org/10.1145/2875501.2875502

[280] Kirk St.Amant. 2019-10-01. The Cultural Context for Communicating Care. *Journal of Technical Writing and Communication* 49, 4 (2019-10-01), 367–382. DOI:http://dx.doi.org/10.1177/0047281619871213 Publisher: SAGE Publications Inc.

[281] Neville Stanton. 2007. Human-Error Identification in Human-Computer Interaction. In *The Human-Computer Interaction Handbook* (2 ed.). CRC Press. Num Pages: 12.

[282] Michael Stead, Paul Coulton, and Joseph Lindley. 2018-06-25. Do-It-Yourself Medical Devices: exploring their potential futures through design fiction. *DRS Biennial Conference Series* (2018-06-25). https://dl.designresearchsociety.org/drs-conference-papers/drs2018/researchpapers/175

[283] Amanda J. Stent, Marie K. Huffman, and Susan E. Brennan. 2008-03-01. Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication* 50, 3 (2008-03-01), 163–178. DOI:http://dx.doi.org/10.1016/j.specom.2007.07.005

[284] David Stewart, Prem Shamdasani, and Dennis Rook. 2007. *Focus Groups*. SAGE Publications, Ltd. DOI:http://dx.doi.org/10.4135/9781412991841

[285] David W. Stewart and Prem N. Shamdasani. 2014-03-20. *Focus Groups: Theory and Practice*. SAGE Publications. Google-Books-ID: 1svuAwAAQBAJ.

[286] Rebecca E. Stewart, Nathaniel Williams, Y. Vivian Byeon, Alison Buttenheim, Sriram Sridharan, Kelly Zentgraf, David T. Jones, Katelin Hoskins, Molly Candon, and Rinad S. Beidas. 2019-06-14. The clinician crowdsourcing challenge: using participatory design to seed implementation strategies. *Implementation Science* 14, 1 (2019-06-14), 63. DOI:http://dx.doi.org/10.1186/s13012-019-0914-2

[287] Allan Stisen, Nervo Verdezoto, Henrik Blunck, Mikkel Baun Kjærgaard, and Kaj Grønbæk. 2016-02-27. Accounting for the Invisible Work of Hospital Orderlies: Designing for Local and Global Coordination. In *Proceedings of the 19th ACM Conference*

*on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. Association for Computing Machinery, 980–992. DOI:http://dx.doi.org/10.1145/2818048.2820006

[288] Christopher Gordon Strachan. 2016-12-01. Design, fiction and the medical humanities. *Medical Humanities* 42, 4 (2016-12-01), e15–e19. DOI:http://dx.doi.org/10.1136/medhum-2016-010934 Publisher: Institute of Medical Ethics Section: Electronic pages - Science fiction and medical humanities.

[289] Lea Strohm, Charisma Hehakaya, Erik R. Ranschaert, Wouter P. C. Boon, and Ellen H. M. Moors. 2020-10-01. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. *European Radiology* 30, 10 (2020-10-01), 5525–5532. DOI:http://dx.doi.org/10.1007/s00330-020-06946-y

[290] Yu-Hsiang Su, Ching-Ping Chao, Ling-Chien Hung, Sheng-Feng Sung, and Pei-Ju Lee. 2020-01. A Natural Language Processing Approach to Automated Highlighting of New Information in Clinical Notes. *Applied Sciences* 10, 8 (2020-01), 2824. DOI:http://dx.doi.org/10.3390/app10082824 Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[291] Lucy A. Suchman. 1987. *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press, USA.

[292] Akiri Surely. 2023-05-18. Applications of Voice User Interfaces in Clinical Settings. (2023-05-18). DOI:http://dx.doi.org/10.48550/arXiv.2305.11914

[293] Theresa Tanenbaum. 2014-09-01. Design fictional interactions: why HCI should care about stories. *interactions* 21 (2014-09-01), 22–23. DOI:http://dx.doi.org/10.1145/2648414

[294] Charlotte Tang and Sheelagh Carpendale. 2009-04-04. A mobile voice communication system in medical setting: love it or hate it?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, 2041–2050. DOI:http://dx.doi.org/10.1145/1518701.1519012

[295] Huimin Tang, Boon Giin Lee, Dave Towey, Kaiyi Chen, Yichu Fang, Runzhou Zhang, and Matthew Pike. 2024. The Audience Effect: Do Observations Change Outcomes in HCI Studies?. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*. 611–620. DOI:http://dx.doi.org/10.1109/COMPSAC61105.2024.00088

[296] Nick Taylor, Jon Rogers, Loraine Clarke, Martin Skelly, Jayne Wallace, Pete Thomas, Babitha George, Romit Raj, Mike Shorter, and Michelle Thorne. 2021-06-28. Prototyping Things: Reflecting on Unreported Objects of Design Research for IoT. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21)*. Association for Computing Machinery, 1807–1816. DOI:http://dx.doi.org/10.1145/3461778.3462037

[297] Selena Taylor. 2022-04-13. Research on Patient safety and interactive medical devices: Realigning work as imagined and work as done. (2022-04-13).

[298] A.M.K. Thomas and A.K. Banerjee. 2013. *The History of Radiology*. OUP Oxford. https://books.google.co.uk/books?id=y44bO1oCho8C

[299] E. J. Thomas, J. B. Sexton, and R. L. Helmreich. 2004-10-01. Translating teamwork behaviours from aviation to healthcare: development of behavioural markers for neonatal resuscitation. *BMJ Quality & Safety* 13 (2004-10-01), i57–i64. Issue suppl 1. DOI:http://dx.doi.org/10.1136/qshc.2004.009811 Publisher: BMJ Publishing Group Ltd Section: Original Article.

[300] Nguyen Minh Truong, Trung Quang Vo, Hien Thi Bich Tran, Hiep Thanh Nguyen, and Van Nu Hanh Pham. 2023-12. Healthcare students' knowledge, attitudes, and perspectives toward artificial intelligence in the southern Vietnam. *Heliyon* 9, 12 (2023-12), e22653. DOI:http://dx.doi.org/10.1016/j.heliyon.2023.e22653

[301] Emmanouil Tsekleves, Andy Darby, Anna Whicher, and Piotr Swiatek. 2017-05-31. Co-designing design fictions:a new approach for debating and priming future healthcare technologies and services. *Archives of Design Research* 30, 2 (2017-05-31), 5–21. https://doi.org/10.15187/adr.2017.05.30.2.5 Num Pages: 17 Number: 2.

[302] Shintaro Tsuji, Andrew Wen, Naoki Takahashi, Hongjian Zhang, Katsuhiko Ogasawara, and Gouqian Jiang. 2021-10-29. Developing a RadLex-Based Named Entity Recognition Tool for Mining Textual Radiology Reports: Development and Performance Evaluation Study. *Journal of Medical Internet Research* 23, 10 (2021-10-29), e25378. DOI:http://dx.doi.org/10.2196/25378

[303] Akshay Vankipuram, Stephen Traub, and Vimla L. Patel. 2018-03-01. A method for the analysis and visualization of clinical workflow in dynamic environments. *Journal of Biomedical Informatics* 79 (2018-03-01), 20–31. DOI:http://dx.doi.org/10.1016/j.jbi.2018.01.007

[304] Mithra Vankipuram, Kanav Kahol, Trevor Cohen, and Vimla L. Patel. 2011-06-01. Toward automated workflow analysis and visualization in clinical environments. *Journal of Biomedical Informatics* 44, 3 (2011-06-01), 432–440. DOI:http://dx.doi.org/10.1016/j.jbi.2010.05.015

[305] Tara Vega, Ariana Bueno, Chad Cerutti, Haoran Chang, Matthew Garvin, Catalina Garza, Parker Kurlander, Nilton Renno, Fernando Figueroa, and Lauren Underwood. Enabling a Voice Management System for Space Applications. https://ntrs.nasa.gov/citations/20210022373 NTRS Author Affiliations: University of Michigan–Ann Arbor, University Space Research Association, Stennis Space Center NTRS Document ID: 20210022373 NTRS Research Center: Stennis Space Center (SSC).

[306] Keith Vertanen. 2006. Speech and Speech Recognition during Dictation Corrections. *Interspeech* (2006).

[307] Johanna Viitanen. 2009-09-01. Redesigning digital dictation for physicians: A user-centred approach. *Health Informatics Journal* 15, 3 (2009-09-01), 179–190. DOI: http://dx.doi.org/10.1177/1460458209337429 Publisher: SAGE Publications Ltd.

[308] Chris Vincent and Ann Blandford. 2011-09-01. Designing for Safety and Usability User-Centered Techniques in Medical Device Design Practice, Vol. 55. 793–797. DOI: http://dx.doi.org/10.1177/1071181311551164

[309] Christopher J. Vincent and Ann Blandford. 2015-02. Usability standards meet scenario-based design: challenges and opportunities. *Journal of Biomedical Informatics* 53 (2015-02), 243–250. DOI:http://dx.doi.org/10.1016/j.jbi.2014.11.008

[310] Shijun Wang and Ronald M. Summers. 2012-07. Machine learning and radiology. *Medical Image Analysis* 16, 5 (2012-07), 933–951. DOI:http://dx.doi.org/10.1016/j.media.2012.02.005

[311] Christopher Watura, Charlotte Kendall, and Paul Sookur. 2022-01-01. Direct Access and Skill Mix Can Reduce Telephone Interruptions and Imaging Wait Times: Improving Radiology Service Effectiveness, Safety and Sustainability. *Current Problems in Diagnostic Radiology* 51, 1 (2022-01-01), 6–11. DOI: http://dx.doi.org/10.1067/j.cpradiol.2021.06.002

[312] Linda Wauben, Johan Lange, and Richard Goossens. 2011-01-01. *Safety lessons from aviation for surgery*.

[313] Robert L. Wears and Marc Berg. 2005-03-09. Computer Technology and Clinical Work: Still Waiting for Godot. *JAMA* 293, 10 (2005-03-09), 1261. DOI:http://dx.doi.org/10.1001/jama.293.10.1261

[314] Lauren H. Williams. 2018. *Interruptions in Radiology: Investigating the Utility of Two Simple Interventions under Differing Rates of Interruption Prevalence*. M.S. https://www.proquest.com/docview/2384584688/abstract/F643BC9E8D244FCDPQ/1 ISBN: 9781392456781.

[315] Lauren H. Williams and Trafton Drew. 2017-02-20. Distraction in diagnostic radiology: How is search through volumetric medical images affected by interruptions? *Cognitive Research: Principles and Implications* 2, 1 (2017-02-20), 12. DOI: http://dx.doi.org/10.1186/s41235-017-0050-y

[316] Rogier Woltjer, Ella Pinska-Chauvin, Tom Laursen, and Billy Josefsson. 2015-09-01. Towards understanding work-as-done in air traffic management safety assessment and design. *Reliability Engineering & System Safety* 141 (2015-09-01), 115–130. DOI: http://dx.doi.org/10.1016/j.ress.2015.03.010

[317] Ming Chao Wong, Elizabeth Cummings, and Paul Turner. 2013. User-centered Design in Clinical Handover: Exploring Post-Implementation Outcomes for Clinicians. *MEDINFO 2013* (2013), 253–257. DOI:http://dx.doi.org/10.3233/978-1-61499-289-9-253 Publisher: IOS Press.

[318] Ming Chao Wong, Paul Turner, and Kwang Chien Yee. Involving Clinicians in the Development of an Electronic Clinical Handover System – Thinking Systems Not Just Technology. (????), 6.

[319] S. H. Wong, H. Al-Hasani, Z. Alam, and A. Alam. 2019-01-01. Artificial intelligence in radiology: how will we be affected? *European Radiology* 29, 1 (2019-01-01), 141–143. DOI:http://dx.doi.org/10.1007/s00330-018-5644-3

[320] Laure Wynants, Luc J M Smits, and Ben Van Calster. 2020-09-09. Demystifying AI in healthcare. *BMJ* (2020-09-09), m3505. DOI:http://dx.doi.org/10.1136/bmj.m3505

[321] Rachel M. Wynn, Jessica L. Howe, Linda C. Kelahan, Allan Fong, Ross W. Filice, and Raj M. Ratwani. 2018-12-01. The Impact of Interruptions on Chest Radiograph Interpretation: Effects on Reading Time and Accuracy. *Academic Radiology* 25, 12 (2018-12-01), 1515–1520. DOI:http://dx.doi.org/10.1016/j.acra.2018.03.016

[322] Emily York, Shannon N. Conley, Anne D. Henriksen, Dorothy Caserta, Noah Etka, Nolan Harrington, Meghan Jennings, Samuel Kodua, Riley Pates, Zachary Sevison, Elizabeth Terry, Sydney VanNostrand, and Karla Vargas. 2019-07. Co-Imagining the Futures of Implementation Precision Medicine Using Scenario Analysis and Design Fiction. *OMICS: A Journal of Integrative Biology* 23, 7 (2019-07), 340–349. DOI:http://dx.doi.org/10.1089/omi.2019.0083 Publisher: Mary Ann Liebert, Inc., publishers.

[323] Brita Ytre-Arne and Hallvard Moe. 2021-07. Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society* 43, 5 (2021-07), 807–824. DOI: http://dx.doi.org/10.1177/0163443720972314