Understanding and Evaluating User Interface Visibility

This thesis is submitted for the degree of Doctor of Philosophy

Ian Michael Hosking
Darwin College

Department of Engineering
University of Cambridge

April 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Ian Michael Hosking
Darwin College, Cambridge
April 2020

Abstract

Understanding and Evaluating User Interface Visibility

Ian Michael Hosking

Technology dominates our lives, mobile technology in particular. In 2016 Apple sold their billionth iPhone. By 2018 they had sold their 2 billionth device based on the same underlying operating system. We access such technology through the user interface (UI) and concerns have been raised about the usability of such devices. The situation has been described by some as a "usability crisis". One of the key issues raised is the lack of visibility of user interface elements, which is deemed to be a critical component of an effective UI.

An initial investigation highlighted that UI visibility can be broken down into three key aspects: Firstly; some user interface elements are effectively **'missing'**; Secondly, they are **'missed'** because they are not seen by the user; and thirdly, they are seen but **'misunderstood'**. Further analysis of the home screen of an iPhone revealed that only **8%** of the available functions were visible at the top level, in other words, **92%** were effectively 'missing'. This raises key questions about how UI visibility can be evaluated, and such evaluation adopted into design practice. This research took a psychophysical perspective to better understand UI visibility. This led to the development of an evaluation framework and associated tool called **vis-UI-lise.** The tool represents UI visibility as a series of 5 hurdles between the user and the interface that have to be overcome for a successful interaction.

This tool was applied to an everyday task on a mobile phone which resulted in highlighting a range of possible usability problems. Comparison of the predicted versus observed problems showed that the vis-UI-lise tool had predicted **74%** of them, a score that compares well with other usability evaluation tools. A training and support package was also developed for the vis-UI-lise tool and evaluated with four different organisations. This provided key insights into how the tool could be improved to fit in with typical design practice. This thesis brings a new perspective to the understanding and evaluation of UI visibility that could have a real impact on the design of everyday user interfaces.

To my family near and far…

and users everywhere who have ‘visibly’ struggled
with the ‘invisible’ elements of user interfaces.

The important design rule of a GUI is visibility [1]

Visibility indicates the mapping between intended actions and actual operations. Visibility indicates crucial distinctions—so that you can tell salt and pepper shakers apart, for example. And visibility of the effects of the operations tells you if the lights have turned on properly, if the projection screen has lowered to the correct height, or if the refrigerator temperature is adjusted correctly. It is lack of visibility that makes so many computer-controlled devices so difficult to operate. And it is an excess of visibility that makes the gadget-ridden, feature-laden modern audio set or video cassette recorder (VCR) so intimidating. [2]

1. Donald A. Norman, “Natural user interfaces are not natural” *interactions* 17:3 (2010):6.

2. Donald A. Norman, *The Psychology of Everyday Things* (New York, NY: Basic Book, 1988),8.

Acknowledgements

An endeavour such as this is built on the foundation of the help, support and guidance of others, for which I am truly grateful. My supervisor, Prof. John Clarkson, encouraged me to tell the story of the research and create a coherent thread from start to finish. My advisor, Dr Nathan Crilly, whose forensic-like ability to constructively critique numerous elements spurred me on to get the detail right. My colleague, Dr Sam Waller, who helped me wrestle with various pieces of software to make the thesis look as it does. My fellow traveller on the road marked PhD, Mike Bradley, who was always the voice of calm encouragement. Anthony Haynes, a writing tutor who made me believe in my ability to write and helped me improve.

At the heart of the research were an array of user participants and usability practitioners who were generous in giving their time and support and who brought the reality of the challenge of user interface visibility to life.

Finally, there was the encouragement of family and friends who cheered me along the journey. David, who diligently read a draft and gave invaluable comments. My brother, who patiently asked and listened to updates on progress. My children and dear wife who always believe in me and give unflinching support.

1. Arnold J. Wilkins, Jennifer Smith, Clare K Willison, Tom Beare, Alexandra Boyd, Gemma Hardy, Louise Mell, Charlotte Peach, and Samantha Harper. "Stripes Within Words Affect Reading," *Perception* 36, no. 12 (2007): 1788–1803. doi:10.1068/p5651.

2. Stephanie Jainta, Wolfgang Jaschinski, and Arnold J. Wilkins. "Periodic Letter Strokes Within a Word Affect Fixation Disparity During Reading," *Journal of Vision* 10, no. 13 (November 1, 2010): 2–2. doi:10.1167/10.13.2

3. Sarah M. Haigh, Nicholas R. Cooper, and Arnold J. Wilkins. "Cortical Excitability and the Shape of the Haemodynamic Response," *NeuroImage* 111 (2015): 379–84. doi:10.1016/j.neuroimage.2015.02.034.

4. "English Pages," Color Universal Design Organization, accessed September 29, 2019, http://www2.cudo.jp/wp/?page_id=1936

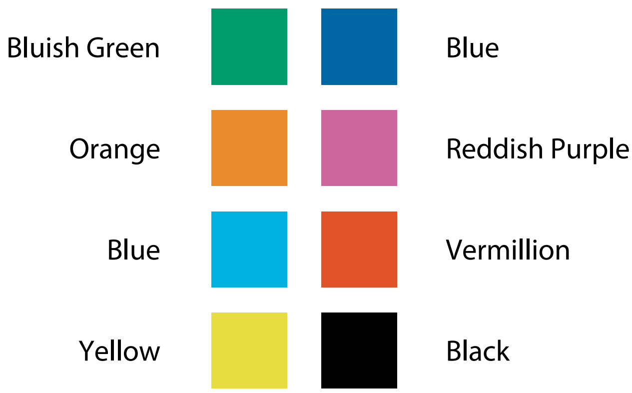
5. Tomohiro Ikeda, Yasuyo G. Ichihara, Natsuki Kojima, Hisaya Tanaka, and Kei Ito. "Color Universal Design: Analysis of Color Category Dependency on Color Vision Type (4)," edited by Reiner Eschbach, Gabriel G Marcu, and Alessandro Rizzi, 8652:86520G–9, SPIE, 2013. doi:10.1117/12.2001582.

6. Masataka Okabe, and Kei Ito. "How to Make Figures and Presentations That Are Friendly to Color Blind People," Color Universal Design Organisation, August 2002. https://jfly.uni-koeln.de/html/manuals/pdf/color_blind.pdf.

Thesis Format

The main font selected for this thesis is 'Verdana'. Verdana has a lower spatial periodicity (stripiness) than other typical fonts [1]. This property makes it easier for the eyes to verge on the letters which can improve reading time and reduce visual stress [1,2]. The impact of this has been shown at a neurological level by the reduction in cortical excitability [3]. Put simply it makes reading easier and reduces the effort required by the brain.

The colour set chosen is that promoted by the Color Universal Design Organization [4]. The set itself is optimised for discriminability across the different colours for different forms of colour blindness [5]. There is a trade-off between set discriminability and contrast when colours are used together e.g. coloured text in a coloured box and the use of white or black text in coloured boxes. This trade-off impacts people with reduced visual acuity (contrast) versus those with colour blindness (set discriminability). This trade-off is mitigated in the various figures in the thesis through the use of larger font sizes where feasible. The colour set [6] is shown below.



Contents

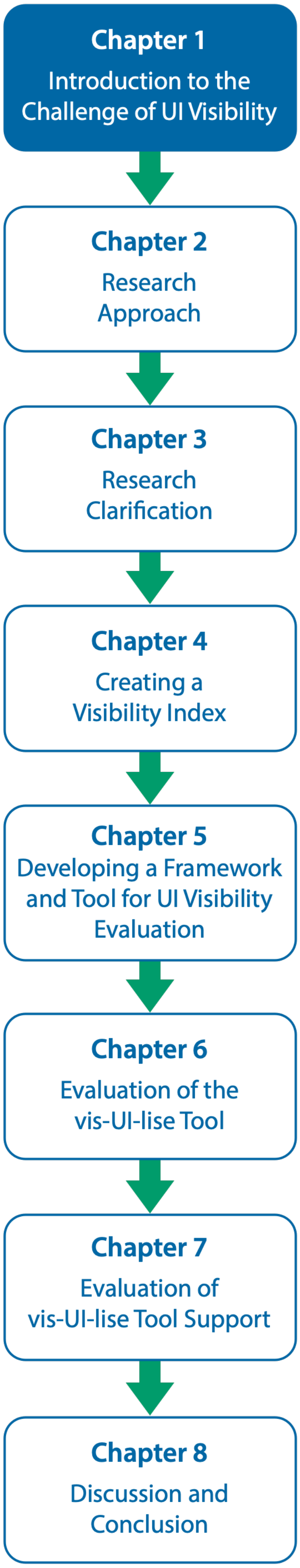
List of Figures

List of Tables

Nomenclature

Term	Definition
DRM	Design Research Method is a research framework for investigating the design process and the development of design tools.
GUI	Graphical User Interface is a general term for user interfaces of electronic devices that have graphical elements over and above a simple text display e.g. using icons. It is closely related to the expression WIMP interfaces (see below).
Guiding questions	High-level research questions that guide this research, that in turn addresses specific research questions.
Hurdle question	A question that forms part of the vis-UI-lise tool that helps to evaluate specific issues of user interface visibility.
IRQ	An Initial Research Question addresses a key issue around user interface visibility.
ORQ	The Overarching Research Question frames the research enquiry from a 'problem within the world' perspective.
PCA	Perception - Control - Action. Is a framework used in the formal evaluation of medical devices to breakdown the interaction between the user and the device.
Seven P's model	A conceptual model of user interface visibility.
Specific research questions	The hierarchy of research questions starting with the ORA which is broken down in the IRQ's and further into the SRQ's.
SRQ	A Supplementary Research Question is a detailed research question that helps in addressing an IRQ.
Three M's Model	Three key user interface visibility problems covering interface elements that are effectively 'missing'; others that are 'missed' because they are not seen by the user; and those seen but 'misunderstood'.
UEM	Usability Evaluation Method is a generic term for a method that systematically evaluates the usability of user interfaces.
uFMEA	User Failure Modes Effects Analysis is the formal process for identifying potential failure modes in a system and their causes and effects.
UI Visibility	User Interface Visibility concerns the quality of a user interface's visual properties.

Term	Definition
Usability practitioner	A professional involved in the design and evaluation of user interfaces.
User participant	A person recruited to assist in the research.
vis-UI-lise	The name given to the tool developed to assist in the evaluation of user interface visibility.
Visibility Index	The percentage of user interface functions that are effectively invisible from a visual perspective to the user.
WIMP	A graphical user interface (GUI) that contains windows, icons, menus and pointers.


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 1

Introduction to the Challenge of UI Visibility

1.1 Personal Motivation

"*Where is the button? it must be there!*". More often than not it was there but for some reason, it was hard to find. It is striking how elusive some functions on the user interfaces of everyday products from photocopiers to desktop software are. What is intriguing is that once the function, often a button, was found it becomes 'obvious'. This personal realisation led to informal observations of others having the same struggle and the issue stood out from anything I had seen before. I have been involved in user interface evaluation and design for over 25 years. During this time a number of projects that I have been involved in have influenced this work. I worked with the National Physical Laboratory on applying the Diagnostic Recorder for Usability Measurement (DRUM) tool to the analysis of a number of systems (Macleod and Rengger, 1993). DRUM assists in a detailed time-based analysis of task performance to generate usability measurements. This required setting up a video-based usability laboratory to record users for subsequent analysis. Linked to this quantitative analysis was the use of questionnaires such as the Software Usability Measurement Inventory (Kirakowski and Corbett, 1993; van Veenendaal, 1998) and the NASA TLX (Hart and Staveland, 1988).

In parallel to this, I was involved in research around the use of task analysis for requirements definition (Ormerod, Richardson, and Shepherd, 2000). Task analysis has been a key habit of mind (Costa and Kallick, 2000) throughout my career, particularly in terms of breaking tasks down to inform requirements capture and drive interaction design. These and other projects included the analysis of large process control systems, customer service systems, medical devices, mobile phones, heating controls and even a smartwatch. I have watched directly and analysed many hours of video recordings of users struggling with interfaces, yet this problem seemed much more complex than many that I have witnessed over the years. Why do user interface controls and elements go ‘missing in action’ as users attempt to complete a task?

Similar problems with visibility hit the headlines from time to time. For example, the Gorilla Experiment, where people are asked to watch a video and count the number of times a ball is passed between a group of people (Simons and Chabris, 1999). During the sequence, a person dressed as a gorilla walks into the middle of the scene, beats their chest and walks off. Despite being centre stage nearly 50% of participants do not 'see' the gorilla. This phenomenon is known as inattentional blindness. More intriguing still is the example from the Himba tribe in Namibia who do not have words to separately categorise blue and green, and when tested on differentiating between the two, they struggle. Conversely, the Russian language has two different words for blue leading to better discrimination in tests with regards to this colour. This leads to researchers arguing that language itself shapes our perception of colour (Roberson, Davidoff, Davies and Shapiro, 2005; Roberson and Hanley, 2007; Winawer et al., 2007).

Missing a gorilla in full view, a tribe struggling to distinguish between the green and blue, and users struggling to find buttons on everyday devices. It all points to there being more to 'visibility' than 'meets the eye' both literally and metaphorically. As a designer is it important to look beyond the graphical image of a user interface to the psychological element to design effective user interfaces? These initial observations and reflections lead to the following overarching research question:

> *How can current approaches to the inclusive design of user interfaces be improved to more effectively address the complexity and variability of human vision?*

1.2 Technological Motivation

Having highlighted a problem, it is important to ascertain whether it is worth addressing, in other words, how much it matters. User interfaces are the means with which we access and control technology. The range and reach of technology are increasing dramatically, as is our reliance on it. The last 100 years have seen a dramatic change in technology and the pace of change is increasing. Figure 1-1 is a graph showing the adoption of a range of technologies (Ritchie and Roser, 2017). The dashed lines highlight technologies that have gained rapid adoption in the last few decades. It took 46 years for electricity to reach 25% of US households but only 7 years for the internet to reach this mark (Kurzweil, 2005).

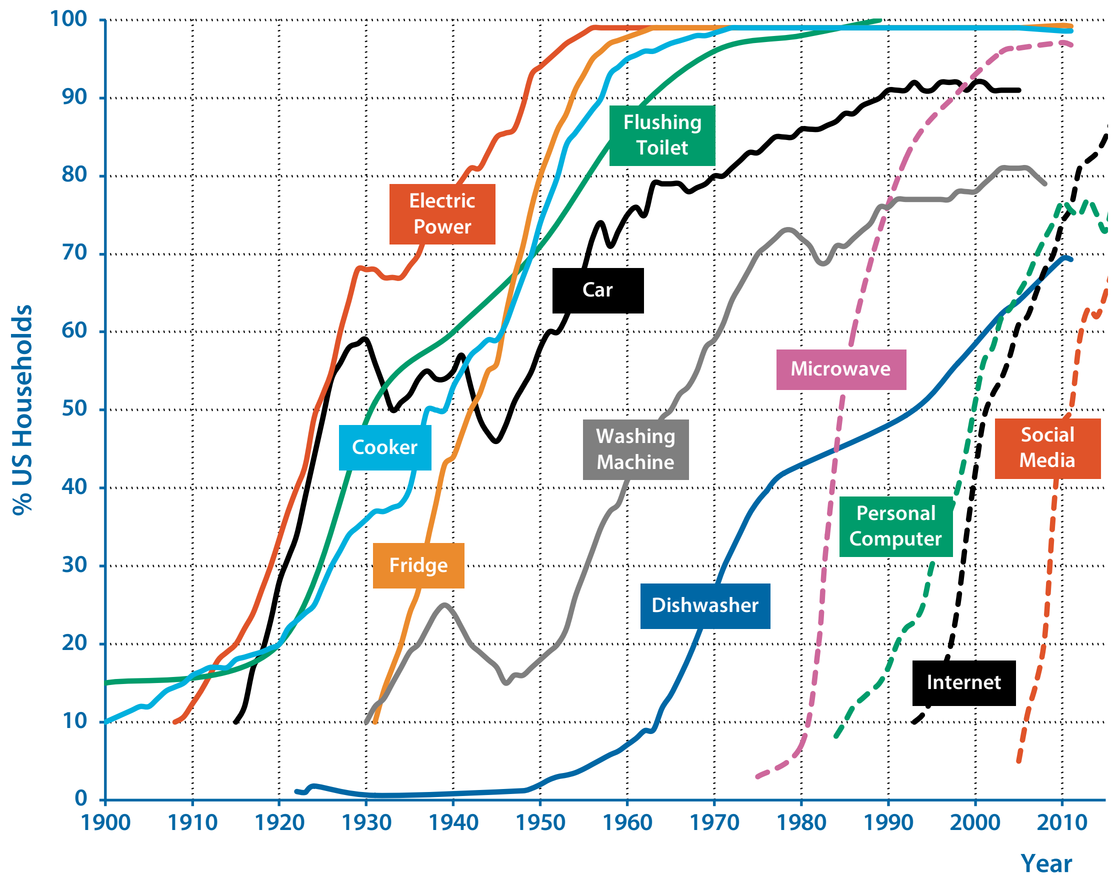


Figure 1-1: The adoption of communications technologies using a selection of data from Ritchie and Roser (2017)

The Internet is a key part of the information and telecommunications age. This has been defined by (Perez, 2010) as the 5th technological revolution (See Table 1-1) and is just a few decades old.

Table 1-1: A categorisation of technological ages from Perez (2010)

Technological revolution	Popular name for the period	Big bang initiating the revolution	Year	Core country or countries
First	The Industrial Revolution	Arkwright's mill opens in Cromford	1771	Britain
Second	Age of Steam and Railways	Test of the Rocket steam engine for the Liverpool–Manchester railway	1829	Britain (spreading to Europe and USA)
Third	Age of Steel, Electricity, and Heavy Engineering	The Carnegie Bessemer steel plant opens in Pittsburgh, PA	1875	USA and Germany forging ahead and overtaking Britain
Fourth	Age of Oil, the Automobile, and Mass Production	First Model-T comes out of the Ford plant in Detroit, MI	1908	USA (with Germany at first vying for world leadership), later spreading to Europe
Fifth	Age of Information, and Telecommunica-tions	The Intel microprocessor is announced in Santa Clara, CA	1971	USA (spreading to Europe and Asia)

This revolution has in part been driven by the increasing availability of computational power combined with digital storage and 2-way communication. The shift to digital storage and the resultant growth in the amount of information stored has been dramatic. Hilbert and López (2011) have estimated the growth in these three key areas. Figure 1-2 shows the dramatic impact of the information age in terms of information storage. It is estimated that, in 1986, only 0.8% of the World's storage was digital. The estimate for 2007 rises to 94%, with the total capacity increasing over 80-fold in 20 years.

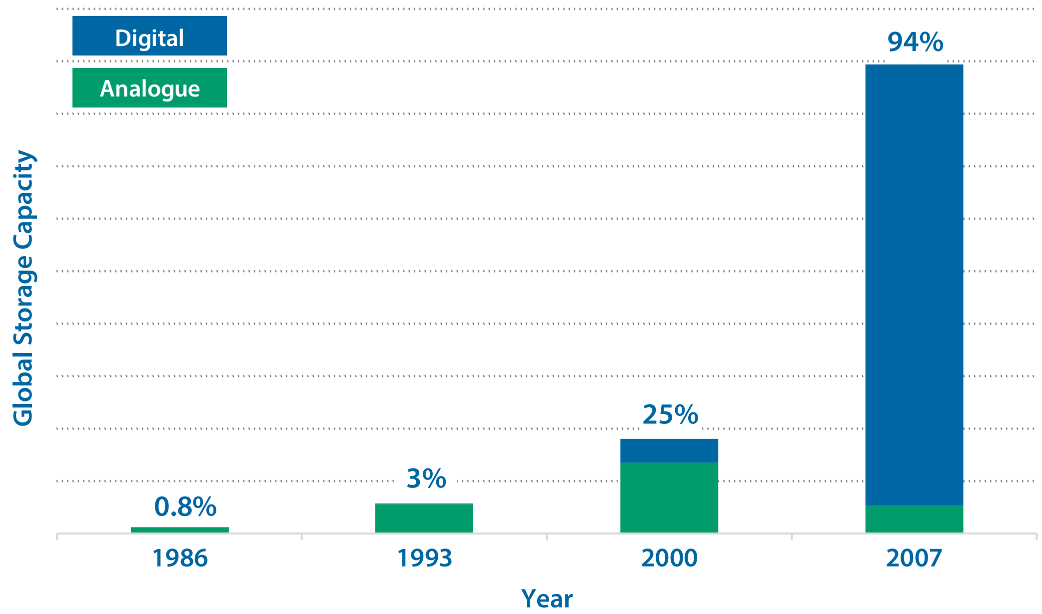


Figure 1-2: The growth in global data storage and the rise of digital storage using data from Hilbert and López (2011)

Within the last 10 years, we have also seen the rise of smartphones and touch devices. Apple alone has been responsible for producing over one billion iPhones (Apple Inc., 2016a). Not only have we seen changes in the underlying technology but also the user interfaces that access this technology. We are in what has been described as a post-WIMP (Windows, Icons, Mice, Pointer) era (Nielsen, 1993a; Gentner and Nielsen, 1996; van Dam, 1997; 2001). The mice and pointer have been replaced by the finger or in some cases just your face which can be used to login to a computer or phone.

These rapid technological developments have led to what has been described by Norman and Nielsen (2010) as a "usability crisis".

At one level it is self-evident that visibility is key. Close your eyes and try and use a computer and you will 'see' how critical the visual elements are. Norman is unequivocal regarding the importance of visibility.

> *"The important design rule of a GUI is visibility" (Norman, 2010, p.6)*
>
> *"Visibility indicates the mapping between intended actions and actual operations. Visibility indicates crucial distinctions—so that you can tell salt and pepper shakers apart, for example. And visibility of the effects of the operations tells you if the lights have turned on properly, if the projection screen has lowered to the correct height, or if the refrigerator temperature is adjusted correctly. It is lack of visibility that makes so many computer-controlled devices so difficult to operate. And it is an excess of visibility that makes the gadget-ridden, feature-laden modern audio set or video cassette recorder (VCR) so intimidating." (Norman, 1988, p.8)*

Despite these robust statements, Norman does not explicitly define ‘visibility’. It is obvious on one hand and complex on the other and something this thesis aims to address. Definitions aside, the scale and pace of change coupled with the difficulties with user interface visibility highlight the importance of this issue.

1.3 Social Motivation

Not only is the technological landscape changing but the world's demographic structure too. The population is ageing. In 1950 there were around 200 million people over 60 (United Nations Department of Economic Affairs Population Division, 2009). By 2050 this number is estimated to exceed 2 billion (Division, 2017). With ageing comes difficulties with adopting new technologies from a sensory, physical and cognitive standpoint (Keates and Clarkson, 2004).

With regard to vision, Wolffson and Davies (Wolffsohn and Davies, 2018) estimate that age related loss of near vision, critical for most user interfaces, impacts over a billion people worldwide and is set to rise with an ageing population. Whilst it is possible to correct for near vision with glasses, it requires the user to have glasses and that they have them to hand. It is estimated that unmanaged presbyopia (normal age-related vision issues) is as high as 50% and 34%, in developing and developed countries respectively, and that this condition impacts task performance.

Ageing, and related vision issues are barriers to realising the benefits of advances in technology. It is not surprising that UK Governments Digital Strategy makes inclusion a priority (UK Digital Strategy). With the population continuing to age, governments are looking to technology to address demands in key areas such as healthcare (National Information Board, 2014).

The user interface represents both a potential hurdle to the adoption of key technologies and an opportunity to make adoption easier. Visibility is a key component in this. The wider social context, outlined here, further strengthens the case for investigating UI visibility.

1.4 Research Motivation

Drawing together the technological and social perspectives we see three key drivers:

1. Visibility is a critical component of most user interfaces
2. We have an increasing reliance on technology, with billions of smart devices in use
3. The global population is ageing and the 'normal' ageing of the eye impacts over a billion people worldwide in the critical area of near vision

The initial user observations, and examples such as the Gorilla Experiment and the Himba tribe, point to needing to understand visibility in terms of the psychophysics of vision. This complexity leads to it being a problem that would benefit from the rigour of academic investigation.

1.5 High-Level Research Approach

As a piece of research, it sits within the widely used definition of research (Coryn, 2006) that was devised by the OECD and published in the Frascati Manual (OECD, 2002). It is used by the University of Cambridge (University of Cambridge, 2019) and represents the broad context for this research in terms of a *"creative work undertaken on a systematic basis in order to increase the stock of knowledge".*

To help further frame the research, ten generic, 'guiding questions' were devised based on Merriam's (2002, p.23) helpful checklist for evaluating qualitative research. These high-level questions provide a 'start to finish' set that guides the research and the content of the thesis. The chapters of the thesis will address these questions hand in hand with the 'specific research questions' directly related to the problem. Indeed, these 'guiding questions' helped in the development of the 'specific research questions'. The guiding questions are as follows:

1. What is the problem?
2. Is it worth solving?
3. Can it be suitably framed as a question suitable for academic enquiry?
4. How well does the literature already answer the question?
5. What is the gap? (in the academic literature)
6. What conceptual understanding is required to study the gap?
7. What is an appropriate empirical approach (study) to address it?
8. What do the results tell us?
9. What is the contribution to the literature?
10. What else needs to be done?

This chapter has started to address the first two questions in the form of outlining the problem and the socio-technical importance of it. The third question, regarding framing the problem as a research question or questions, is what drives the development of the 'specific research questions'. It was something that evolved through the research and is structured using an overarching research question (ORQ) that is further broken down into three initial research questions (IRQ) as follows:

The overarching research question (ORQ) is:

> *How can current approaches to the inclusive design of user interfaces be improved to more effectively address the complexity and variability of human vision?*

This leads to an initial set of research questions (IRQ's) as follows:

> **IRQ1:** *What is UI visibility?*
>
> **IRQ2***: What problems is it causing users?*
>
> **IRQ3:** *What can be done to improve it?*

The research is situated within the area of inclusive design as there is a particular emphasis on addressing diversity in the population, in particular concerning the ageing eye and cognitive demands of complex interfaces. Linked to this is the role of design in developing user interfaces and therefore the initial aim is to help usability practitioners understand user interface visibility and to help them address it within the design process. The overall framework for the research is based on the Design Research Method (DRM) (Blessing and Chakrabarti, 2009) which will be discussed in detail in Chapter 2.

1.6 Outline of Thesis

1.6.1 Structure

The thesis is broken down into the following chapters and appendices which are shown in Table 1-2. It is worth noting that Chapter 3, 'Research Clarification', is an expression used in the DRM approach and equates to the literature review.

Table 1-2: Thesis outline by chapter

No.	Title
1	Introduction to the Challenge of UI visibility
2	Research Approach
3	Research Clarification
4	Creating a Visibility Index
5	Developing a Framework and Tool for UI Visibility Evaluation
6	Evaluation of the vis-UI-lise Tool
7	Evaluation of vis-UI-lise Tool Support
8	Discussion and Conclusion
	Bibliography
A	Evaluation of Windows Phone 10 Scenario using vis-UI-lise
B	vis-UI-lise Training Presentation
C	vis-UI-lise Evaluator Guide
D	vis-UI-lise Quick Start Guide
E	vis-UI-lise Evaluation Template
F	vis-UI-lise Practice Exercise
G	Selected Interview Responses Regarding Key vis-UI-lise Terms
H	Modified Technology Familiarity Questionnaire

1.6.2 Format

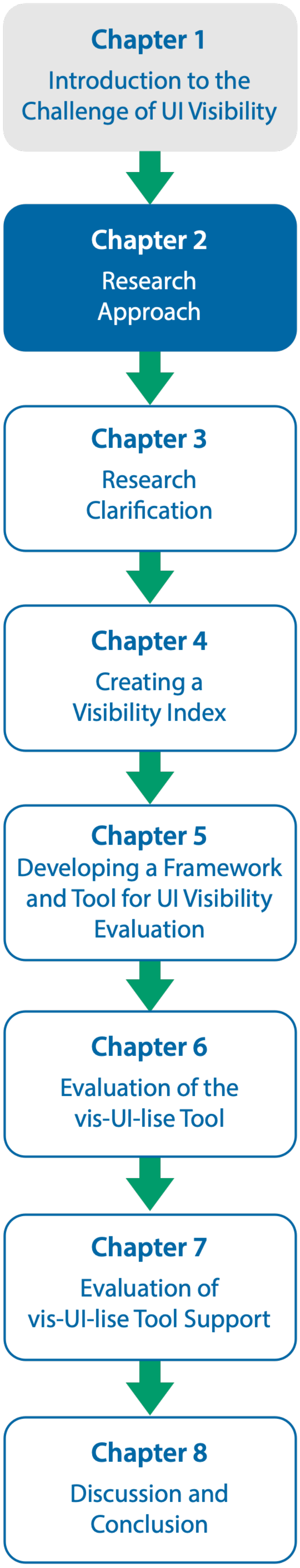
The thesis is formatted with the following features to aid navigation and to maintain an overall context for each chapter:

- Chapter headings are a double-page spread with the chapter number and title on the right side and on the left a diagram of where it fits within the overall thesis
- Each page header contains the chapter title only to make it easy to find chapters
- Figure and table numbers are preceded by the chapter number e.g. Figure 1-1 is the first figure within chapter one.

1.7 Conclusion

User interface visibility, or the lack of it, represents a problem that impacts billions of people on a daily basis. Furthermore, there is more to visibility than meets the eye both literally and metaphorically. This complexity is seen in examples, such as language shaping our perceptual categorisation of colours, and issues of inattentional blindness, even when the thing that is plain sight is a gorilla. The importance of visibility in user interfaces at one level is obvious, one has to simply close your eyes and try and interact with a device. This importance is also clearly stated by academics in the field, yet what is less evident is a robust understanding of what UI visibility is and how to evaluate it. This issue forms the basis for the overarching research question and three further initial research questions aimed at unpacking it.

With regard to the broader ten guiding questions, the first three have been addressed in part. Namely, a problem has been identified (1), the significance of it outlined (2), and the problem framed as an initial set of research questions (3). To answer all the questions raised requires the selection of an appropriate research methodology to answer them in a rigorous manner. This is the focus of the next chapter.


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 2

Research Approach

2.1 Introduction

Initial observations and analysis have established a potential problem with UI visibility. Its significance in terms of the billions of people it can impact has been outlined. Finally, the problem has been framed in terms of a research question. The overarching research question (ORQ) is:

> *"How can current approaches to the inclusive design of user interfaces be improved to more effectively address the complexity and variability of human vision?"*

This chapter will outline the research approach to address this question and the three initial research questions (IRQs) described in the previous chapter.

2.2 Research Approach

To describe the overall research approach a simple model of the respective elements was developed based on the work of Guba (1990), Crotty (1998), Creswell (2009), Greene (2006), Rocco & Plakhotnik (2009) and Toulmin (1958). This model is shown in Figure 2-1.

Figure 2-1: A simple model for the key elements of the research approach

The aspects of these elements, specific to this research, are described in the following sections.

2.2.1 Research Stance

Researching complex phenomena can rarely be reduced down to a simple set of experiments. Even when experiments are possible and valid, it can be hard to generalise these into different situations. Not unsurprisingly this complexity has led to a range of views and approaches to research. Understanding these and stating the particular philosophical stance taken is important to position the research and to make transparent to the reader the perspective taken.

A philosophical stance is either explicitly or implicitly linked to a research paradigm. Guba (1990) describes a paradigm as a *"basic belief system"* that guides our actions, which can cover a broad range of approaches to areas such as social, legal and religious action. As he points out, the focus for academic research is *"those paradigms that guide disciplined inquiry"*. It is worth noting that Guba acknowledges the difficulty of defining the term 'paradigm' that was popularised in academic circles (see Guba, 1990; Crotty, 1998) by Kuhn (Kuhn, 1962).

Guba states that different paradigms can be characterised in terms of three basic questions as follows:

1. Ontological: What is the 'nature' of the 'knowable'?

2. Epistemological: What is the relationship between the 'knowable' and the 'knower'?

3. Methodological: How can the 'knower' find out about the 'knowable'

He points out that *"...paradigms are human constructions, and hence subject to all the errors foibles that inevitably accompany human endeavors"*. He then summarises different paradigms based on these three questions. These are collated in Table 2-1 below:

Table 2-1: Hinrichs' (2010) summary of Guba's (1990, pp.17-27) description of 4 paradigms based on their different ontological, epistemological and methodological perspectives

	Positivism	Post-Positivism	Critical Theory	Constructivism
Ontological	Realist	Critical Realist	Critical Realist	Relativist
Epistemological	Dualist/ objectivist	Modified objectivism	Subjectivist	Subjectivist
Methodological	Experimental/ manipulative	Modified experimental/ manipulative	Dialogic Transformative	Hermeneutic Dialectic

Another framework for research paradigms is provided by Crotty (1998) that shows a hierarchical view of a research paradigm (See Figure 2-2). Crotty takes the view that ontological and epistemological views emerge together and his framework excludes ontology, however, he says that if it was in the framework then it would sit alongside epistemology. Figure 2-2 includes this modification to help relate it to the work of Guba. Part of the value of Crotty's view is to bring in the distinction between methodology and method. It also brings in the notion of 'theoretical perspective' and it could be viewed that Crotty's use of this combined with his view of epistemology broadly map to Guba's use of paradigm. The theoretical perspective is key in selecting an appropriate methodology.

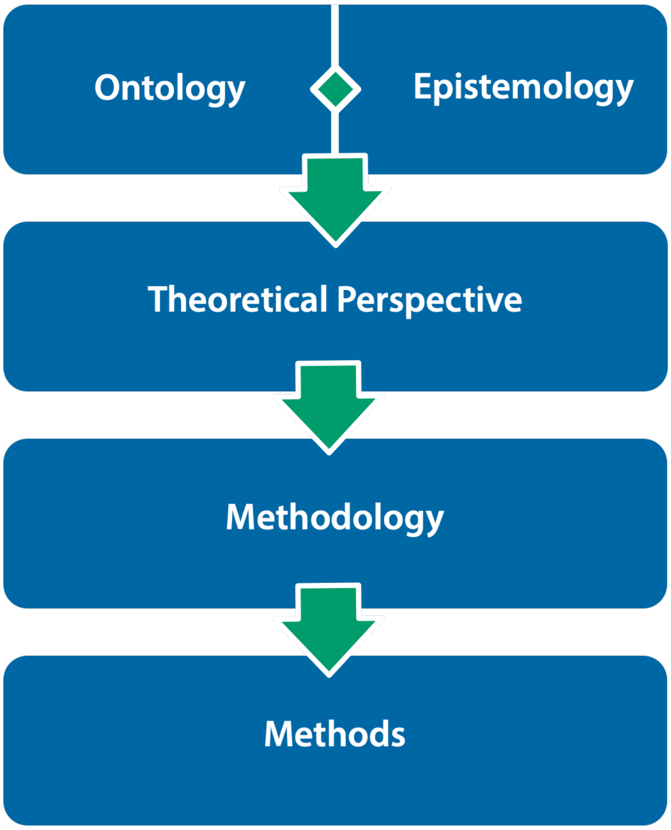


Figure 2-2: A modified representation of Crotty's (1998) research schema showing the position and interaction of epistemology and ontology

This brief discussion represents a small snapshot of a complex, controversial and evolving area. However, it provides a context to state the research stance of the author. The author studied physics as an undergraduate and on reflection resulted in a broadly positivist stance. Working in the field of design, with a focus on usability, has led to a progression to a post-positivist, critical realist stance as defined by Guba (1990), while acknowledging that there is value in different stances. Guba (1990, p.20) neatly describes the critical realist position as one that acknowledges that there is a *"real world"* driven by *"real-world causes"* that exists but that humans' ability to see this is hampered by their imperfect sensory and cognitive abilities. Blumer (1969, p.25) argues that the way we view the world, our picture of it, strongly influences the scientific study of it. He puts this succinctly and forcefully when he states that it has a *"fundamental and pervasive effect wielded on the entire act of scientific* *inquiry by the initiating picture of the empirical world, it is ridiculous to ignore this picture"*. Understanding the underlying basis and the premises is a key part of the research approach. Also, Wilson (1983) argues that one picture is not enough and that it requires taking multiple perspectives to access the underlying truth.

The research stance sits at the centre of Figure 2-1 as it influences and is influenced by the key elements of a research approach. Having stated the particular philosophical stance taken, linked to the specific research paradigm of critical realism, it is important to understand the research was undertaken within the Engineering Design Centre (EDC) at the University of Cambridge. As such this represents a community of practice which Lave & Wenger (1991) argue is an *"intrinsic condition"* for the existence of knowledge and the support of the interpretation of it. This represents what has been described as a social epistemology (Egan and Shera, 1952; Goldman and Blanchard, 2018), which places the researcher and research into a social context. As a community of practice, the EDC works in the fields of engineering design, inclusive design and human-computer interaction, all of which will be described in the next section.

A critical realist stance leads to valuing both quantitative and qualitative approaches and potentially, more importantly, the different perspectives across different disciplines. In particular, the differences between the broad areas of engineering, psychology and social sciences are seen as constructive in offering different insights to the underlying reality. Also, this stance brings appropriate caution to the interpretation of results and how they can be generalised.

2.2.2 Research Field

Within the broad disciplines mentioned, this research draws from three more specific complementary disciplines. These are outlined below highlighting key aspects for this research as follows.

Engineering Design

Engineering design was framed by Dixon (1966) as the intersection of engineering and societal axes. This resonates strongly with the socio-technical description the research problem described in Chapter 1. This is shown in Figure 2-3 below.

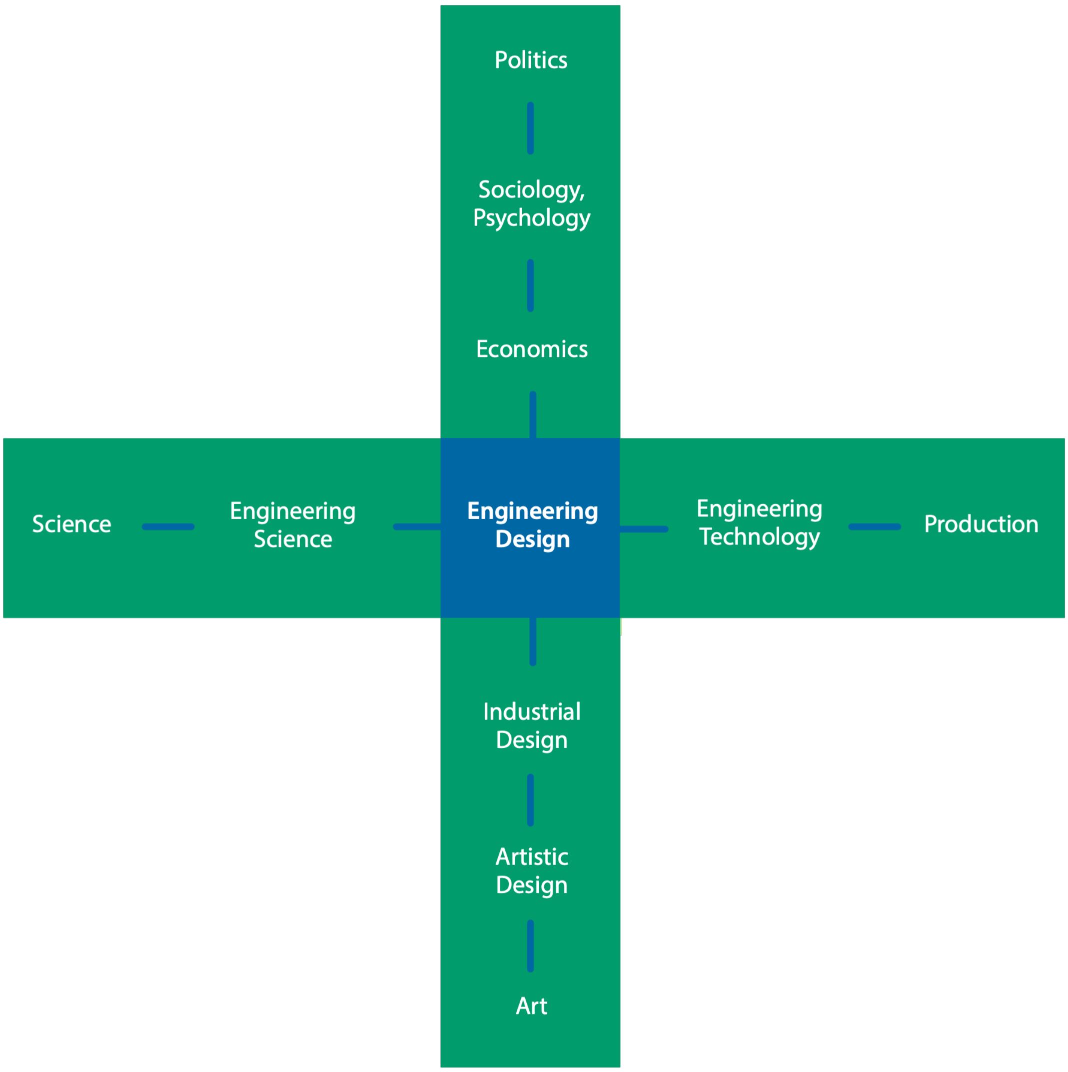


Figure 2-3: Engineering design as the intersection between engineering and social axes based on Dixon (1966)

The definition of engineering design has been refined further by Dixon (1987) and others such as Penny (1970) and Pahl and Beitz (1988). A key aspect of this definition for this enquiry is the emphasis on a process that embodies a systems approach. Systems thinking represents a core habit of mind (Costa and Kallick, 2000) for both the researcher and the research group and further influences the overall research stance.

Human-Computer Interaction (HCI)

The focus on specific aspects of user interfaces leads to drawing on a variety of elements from the field of HCI. The exact nature and make up of this field is contested (Myers, 1998; Grudin, 2012; Karray, Alemzadeh and Saleh, 2008; Carroll, 2010). Indeed, Carroll has described it as a *"meta-discipline"* with its boundaries growing over time. Figure 2-4 is a modified version of a multidisciplinary view from Rogers, Sharp, and Preece (2011). It provides a useful categorisation of academic disciplines, interdisciplinary fields and design practice. Inclusive design, vision science and animation have been added to show how they relate to the field as a whole. These are highlighted in Figure 2-4 with coloured borders. As the field has evolved the emphasis has shifted from 'interaction' to 'experience' (Wright and McCarthy, 2010). The nuances of this debate are not directly relevant to this research and therefore the terms are treated as interchangeable. What matters is accessing the most relevant literature and being able to position any contribution within it. Also, of note is the multidisciplinary nature echoing with that outlined for the field of engineering design, as shown in Figure 2-3, and as such is complementary to it.

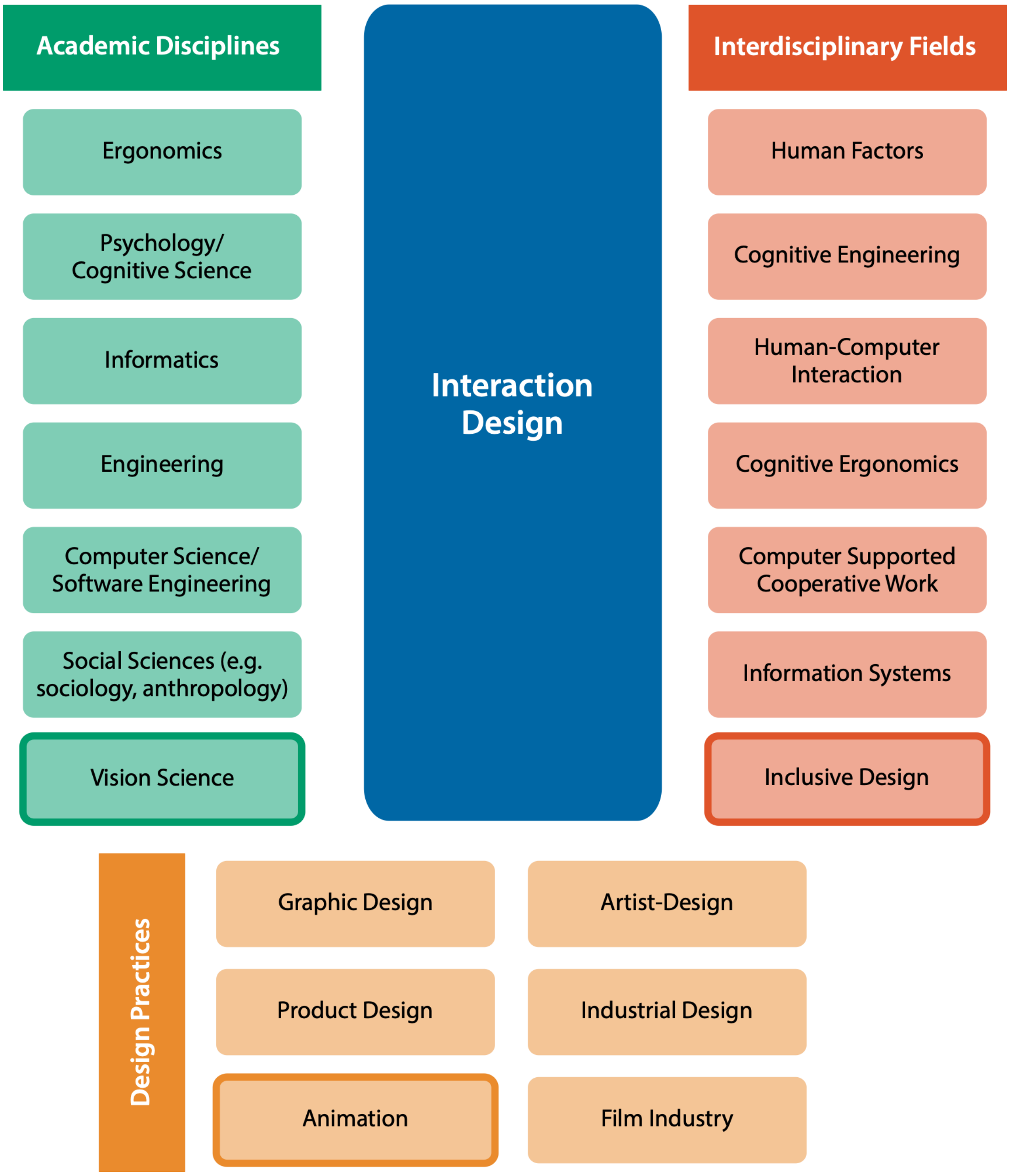


Figure 2-4: An overview of the multidisciplinary nature of interaction design based on Rogers, Sharp and Preece (2011, p.10) with the additional 'fields' highlighted with coloured borders

Inclusive Design

Inclusive design is a response to the need to design for diversity in the population that occurs through issues such as congenital disability and age-related functional loss, but can be extended to other issues such as gender (Hosking, Waller and Clarkson, 2010; Newell et al., 2010; Waller, Bradley, Hosking and Clarkson, 2013). A key aspect for this research is the impact of ageing, in particular, it is reasonable to assume that older people will be impacted more by poor UI visibility. This impact is due to age-related reduction in visual acuity and cognition impacting a user's psychophysical visual performance, which is critical in being able to 'see' elements of a user interface.

It is worth noting that 'inclusive design' is also called 'design for all' and 'universal design'. The differences between these are debated (Persson, Åhman, Yngling and Gulliksen, 2015) but these do not impact this research and are therefore are deemed to be interchangeable in this case.

Interdisciplinary View

An interdisciplinary view is provided by Shneiderman (2000; 2003) who coined the term 'universal usability' which draws upon the fields of HCI and universal (inclusive) design. This approach aims to push for designs that are more widely accessible. Shneiderman (2000) has a very specific definition which is:

> *"Universal usability can be defined as having more than 90% of all households as successful users of information and communications services at least once a week."*

Although this could be seen as too narrow it does show the intent to address diversity in the population. Also, of note is that Shneiderman highlights a research agenda based around three key areas of: (1) technology variety; (2) user diversity; and (3) gaps in user knowledge.

These resonate with the socio-technical perspective discussed in Chapter 1. Also, this specific interdisciplinary view adds weight to the argument for working across the disciplines outlined.

Stember (1991) outlines three core areas regarding the general case for interdisciplinarity which are:

1. **The intellectual argument:** that concepts, theories and methods from one discipline can enrich another and also that it addresses the fragmentation of different disciplines and the necessary narrowness that specialisation can bring.
2. **The practical argument:** that the world's problems do not fit neatly within the boundaries of a specific academic discipline.
3. **The paedogogical argument:** that learning is held back by disciplinary fragmentation.

Therefore, the general case for interdisciplinarity and the specific example of universal usability help build a justification for drawing on different disciplines for this research, which is also consistent with taking a critical realist stance.

2.2.3 Conceptual & Theoretical Frameworks

The literature is used as a basis for developing conceptual and theoretical frameworks to underpin and position the research. Based on the work of Rocco and Plakhotnik (2009) the two are defined as follows:

- A **conceptual framework** is initially a synthesis from the literature of all the key elements for the research and the relationship between them. This helps position and inform the research and the gaps in understanding. As such the research will help develop the conceptual framework further. In this case, the conceptual framework consists of several diagrammatic models. A conceptual framework is likely to contain existing theoretical frameworks. This is defined below.

- A **theoretical framework** consists of the elements and their relationship to each other that is required to test a theory. This may draw on and contribute to the conceptual framework but has a specific purpose related to testing, whether it is qualitative or quantitative in nature.

This research consists of both with the former leading to the later as the nature of user interface visibility is explored.

2.2.4 Research Framework, Methods & Empirical Research

The research stance and interdisciplinarity informs the selection of the framework for the research. The use of the word 'framework' here is akin to Crotty's (1998) use of the word 'methodology' as shown in Figure 2-2. 'Framework' is chosen for the simple reason of the inherent confusion between the use of the word 'methodology' and 'method'.

This specific framework chosen is the Design Research Method (DRM) (Blessing, Chakrabarti and Wallace, 1998) providing an overarching structure for the research. The DRM was developed to address a lack of rigour in design research and the use of such research in practice (Blessing and Chakrabarti, 2009, p.6). As a framework, it allows a range of research methods to be used within it and acknowledges that the endeavour is inherently iterative. The structure is shown in Figure 2-5 with arrows showing the overall flow of the process and the broad scope for iteration with 'backwards' arrows.

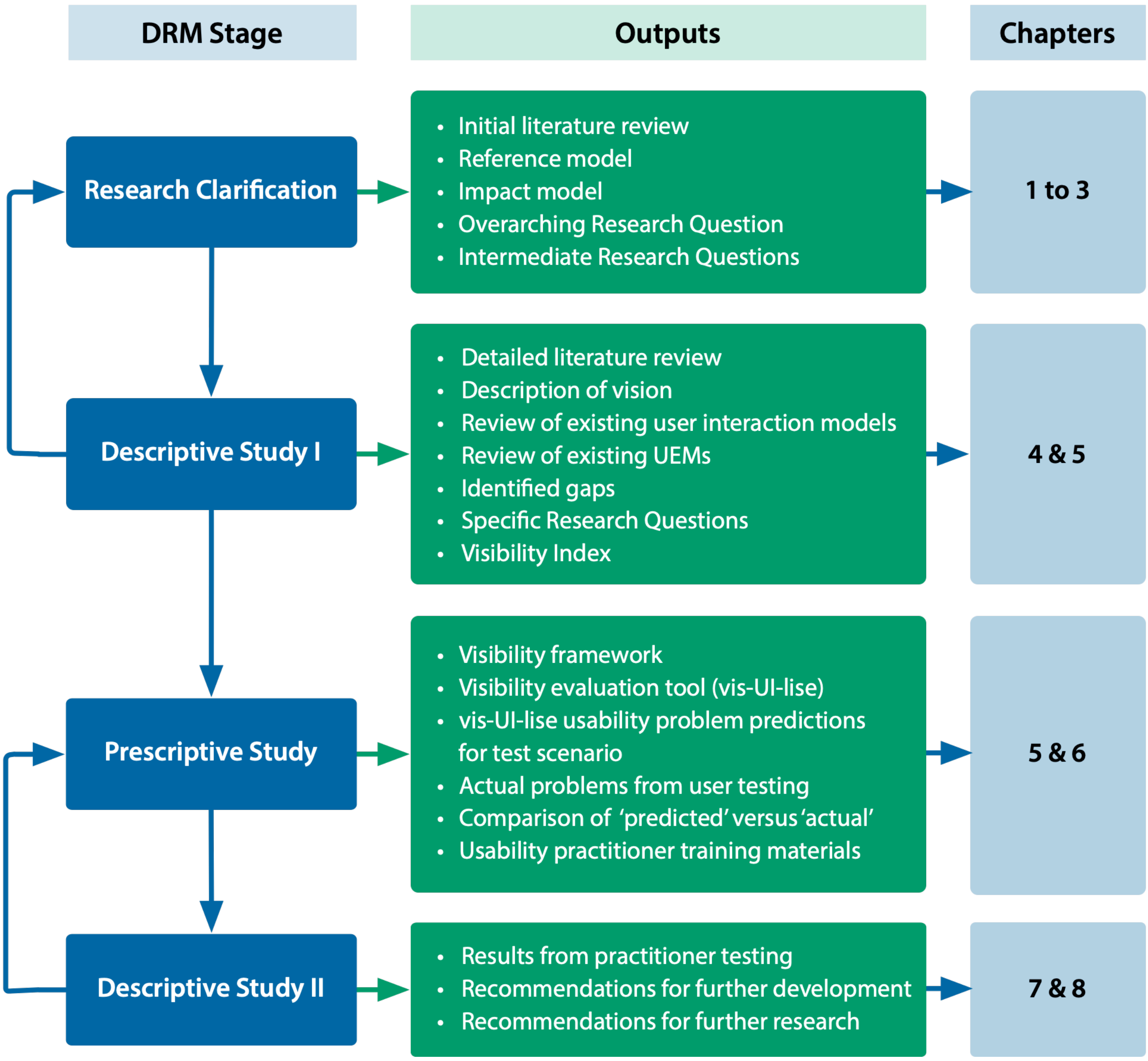


Figure 2-5: Stages of the DRM mapped to key activities and outputs

2.3 Literature Review Approach

The bedrock of a literature review (Cooper, 1985; Boote and Beile, 2005; Schryen, Wagner and Benlian, 2015) is determining if the research question has already been answered and if not why not and therefore to establish gaps in the research. It also informs the development of conceptual and theoretical frameworks as well as the selection and execution of methods. Finally, it also helps position any contribution that is made from the research undertaken. There are various approaches to literature reviews

(Schryen, Wagner and Benlian, 2015) which can be classified using Cooper's taxonomy of literature reviews (1985).

Cooper devised six taxonomic criteria: focus, goals, perspective, coverage, organisation, and audience. These provide a useful structure for not only classifying different types of review but for defining and describing them in the first place. There is potential confusion and overlap between focus and goals and these are merged to just being goals, as the split for this research did not provide any additional clarity. In addition, the types of contribution described by Schryen, Wagner and Benlian (2015), which draws on Cooper's seminal work among others, were used to help structure the thinking around goals. These are shown in Table 2-2 below.

Table 2-2: Different types of contribution from literature reviews described by Schryen, Wagner and Benlian (2015)

Type	Description
Synthesis	Brings together the relevant literature from the domain in a form that underpins and helps shape the research (this is typically part of all academic research).
Adopting a new perspective	A new perspective happens by moving from synthesis to interpretation to produce novel perspectives and insights.
Theory building	Theories can be built by adapting existing ones, synthesizing multiple theories or by building a new one.
Theory testing	From a literature review perspective, theory testing is based on using existing empirical evidence for example through a meta-analysis.
Identification of research gaps	While synthesis is concerned with what is already known, identifying gaps is concerned with what needs to be done to advance the field.
Providing a research agenda	From the research gaps it is possible to propose a research agenda which is distinct from conducting the research itself.

Using this modified form of Cooper's (1985) taxonomic criteria the literature review for this research is described as follows.

Goals

The initial goals of the review are as follows:

1. Establish an underpinning conceptual framework through a synthesis of the literature
2. Highlight any key theoretical frameworks that support the broader conceptual understanding
3. Establish any gaps in understanding or design support
4. Determine appropriate methods for exploring issues and testing any intervention
5. Help determine the positioning of any potential contribution

Perspective

Cooper describes the perspectives as either 'neutral and dispassionate representation' or 'advocacy or a position or perspective'. It also states that this can be seen as a continuum between the two opposites. As desirable as the 'neutral' position is or seems, it would be naïve to assume that this is the stance of the researcher. As the research progresses to the development of an intervention then it is natural for the researcher to be invested in its success and therefore the perspective needs to be checked to ensure there is transparency about the role of the researcher. Wilson (1983, pp. 9,43) describes the role of the researcher as 'detached spectator' and like an artist painting a picture from the literature. As opposed to being an 'actor' who is part of what is happening. DeWalt & DeWalt (2011, pp.22-25) unpack this continuum in more detail. This is summarised in Table 2-3.

Table 2-3: Continuum of researcher interaction from non-participation to full based on DeWalt & DeWalt (2011, pp.22-25)

Category	Membership Role	Description
Non-participation	No role	The researcher is not physically there.
Passive	No role	The research is there but does not interact with the people. They are a 'detached spectator'.
Moderate	Peripheral	The researcher is identifiable to the people but is not actively participating and only occasionally interacts with the people.
Active	Active	The researcher engages with almost everything the people are doing in order to understand it.
Complete	Full	The researcher becomes a full member of the group and as a temporary thing suspends other roles to allow a fuller integration but continues to record observations.
Pure	Full	Where the researcher has 'gone native' and 'sheds' the role of the researcher and has no analytic interest.

For this type of research, one starts as passively observing people struggling with user interfaces. The engagement increases with interaction with both users and designers moving to moderate and occasionally active levels. There is an inevitable risk of losing neutrality in order to gain insight. The key seems to be an awareness, as opposed to denying this risk, to ensure evidence produced is seen within this context. With the evolution of the research, it moves from 'passive' to 'moderate', particularly with the iterative development of an intervention with designers.

Coverage

Cooper (1985) highlights four types of coverage as follows:

1. Exhaustive and comprehensive review and synthesis of the literature
2. Exhaustive but selective use of papers
3. Representative of the topic
4. Central/pivotal papers in the topic

All of these require some inclusion and/or exclusion criteria for whether a paper is relevant to the topic and if it is representative or pivotal within it. Ultimately to be totally confident that a topic has been exhausted with regards to the literature would require all papers in existence to be reviewed for potential inclusion. Such a scope is unrealistic and therefore pragmatism and judgement are required (Boote and Beile, 2006).

As with ‘goals’, Cooper states that the types are not mutually exclusive and for this research, coverage is based around representation and whether the pivotal work is covered. With regard to whether the problem raised has already been directly addressed then a push towards an exhaustive view is made. The process for doing this will be discussed after considering the ‘organization’ and ‘audience’.

Organisation

Cooper (1985) states that reviews can be arranged according to the following criteria: chronological, conceptual and methodological. He also points out that reviews can combine two or more of these. For this research, the predominant form is conceptual using an Areas of Relevance and Contribution (ARC) diagram from the DRM (Blessing and Chakrabarti, 2009, pp.63-66). An ARC diagram helps show not only the contributory areas but whether they are essential or useful, as well as areas of potential contribution.

Audience

PhD requirements aside, the audience is primarily researchers in the fields of HCI and inclusive design but with an important secondary audience of usability practitioners. The multidisciplinary nature of the work and the audience means that the assumption is made that not all readers will be conversant with the concepts and details of the contributory fields. Therefore, a higher level of explanation will be given compared to research conducted within a more constrained field of enquiry.

The Process

Cooper's structure is a useful way of describing the overall nature of the literature review. In addition to this flexible form of criteria-based description there are a number of specific types of literature review that have emerged since Cooper's work (Grant and Booth, 2009). Of these, the integrative review (Whittemore and Knafl, 2005; Russell, 2005) is the most appropriate as it allows for the integration from multiple fields and types (qualitative and quantitative). This contrasts with a systematic review that is appropriate for addressing a well-defined and focussed question where that is a large body of associated literature. This is less suited to this open exploration of the problem.

Another approach from the healthcare domain is a realist review (Pawson, Greenhalgh, Harvey and Walshe, 2005). Although described as a systematic approach it is akin to an integrative review and is aimed at addressing the complexities and ambiguities of real-world systems. Drawing on these different approaches helps define a specific process for conducting the review. This is outlined in Figure 2-6.

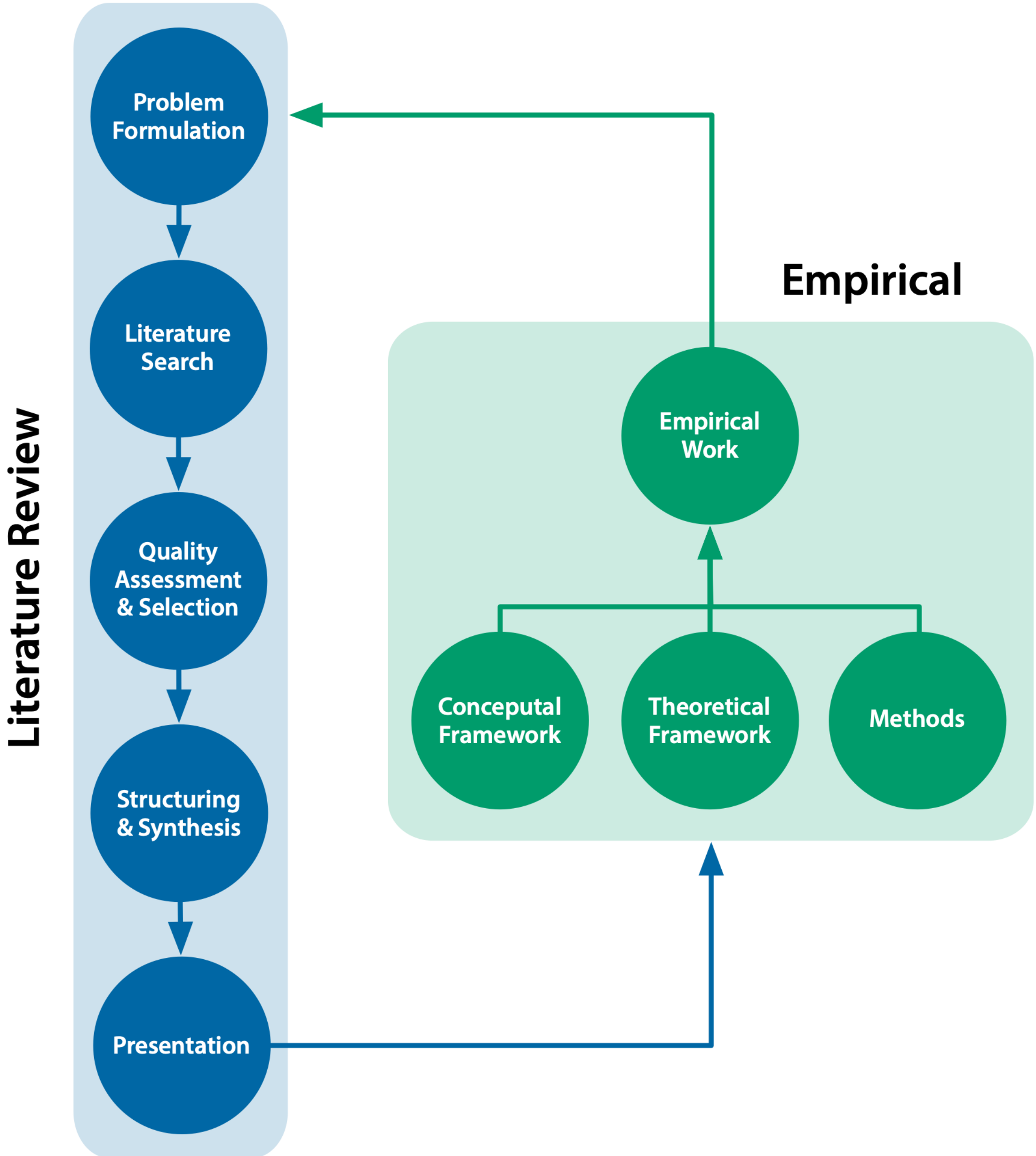


Figure 2-6: Literature review process based on Whittemore and Knafl (2005); Russell (2005); Pawson et al. (2005); and Rocco and Plakhotnik (2009)

This process is highly iterative with a strong dynamic interaction between the empirical work and the literature review. However, the diagram gives a good overview of the key activities and their broad sequence. These are described in more detail as follows.

Problem Formulation

The problem formulation starts with the motivation (Chapter 1) and is refined through the Overarching Research Question (ORQ) and its associated Initial Research Questions (IRQs).

The Literature Search

The literature search was conducted using four key sources as follows:

- A variety of academic search engines
- Non-academic sources such as content produced by non-academic practitioners, news articles or simple Internet searches
- Bibliographies from academic and grey literature
- Discussion with practitioners in the field

Searching was a combination of 'bidirectional citation' (earlier and later related citations) and 'keyword' searches. The bidirectional citation searches were seeded by seminal or key papers which Hinde and Spackman (2014) describe as *"initial pearls"*. This is the basis for what they describe as a *"snowballing"* approach to uncover the key literature in an area. Such an approach overcomes the weaknesses of keyword searching with Boolean logic, such as missing relevant papers where terminology has changed over time.

Quality Assessment & Selection

Papers that are identified through searching require quality assessment. This was based on a range of criteria but is ultimately a human judgement (Wilson, 1983; Tseng and Fogg, 1999; Rieh, 2002; Liu, 2004). The criteria that contributed to the judgment included the following:

- The source e.g. was it from a peer-reviewed journal and if so, what is its impact factor
- The author(s), their institution(s) and their wider body of research

- The bibliography in particular if it cites the seminal work in the field
- The methodology for the research
- The relationship to other work and in particular to seminal work in the field and its level of alignment with it
- Citation by other key authors and overall citation counts

A stance was taken to look beyond reputational factors (Liu, 2004) to the quality of the work as described and how it related to other work. Bibliometrics are potentially useful but high citation counts are hard to interpret, for example, are people citing because they agree or disagree with the work or simply because everyone else does.

Quality assessment goes hand in hand with relevance to the concept being considered and what it adds to the emergent conceptual framework(s). Ultimately it is a human judgement that is informed by the properties outlined.

Structuring and Synthesis

A basic structure was provided by the use of an ARC Diagram as discussed earlier. This was augmented with the development of multiple conceptual frameworks. These interact as the development of the conceptual frameworks drives additional searching of the literature. Synthesis primarily occurs through the connecting of related literature that produces further clarification and insight.

Presentation

'Presentation' occurred in a number of forms. The initial one was the ARC diagram that was annotated with the 'initial pearl' papers that were either seminal work for different aspects of the research or led to finding them (this is shown in next Chapter in Figures 3-12 and 3-13). The development of various frameworks and models helped significantly to further structure the literature and guide further exploration. Finally, during the final writing process, where helpful tables of papers for each section where produced to summarise the key contributions from each paper.

2.3.1 Argumentation - Making Rigorous Claims

The core output of the research approach is a body of evidence about the nature of the problem and potential solutions that might address it. Drawing conclusions from this body of evidence requires equal rigour to that which produced it in the first place. As has already been stated the approach is aimed at addressing the complexity of the real world and therefore an approach to argumentation is required to match this.

Argumentation has developed from the formal logic of the Greeks to complementary modern forms of informal logic (Johnson, 2014; Groarke, 2017). Johnson (2014, p.29) summarises the view that the certainty of deductive logic is rare in everyday life and therefore an approach is required that deals with the uncertainty and nuances of an evidence base. Informal logic approaches aim to address this need and a number of approaches have been developed (Johnson, 2014, pp.1-14). The Toulmin model has been chosen as it is well established (Toulmin, 1958; Wood, 2004; Booth, Colomb and Williams, 2008; Johnson, 2014). As with all models, there are issues and in particular, the notion of 'warrants' can cause confusion. In the forward to the third edition of The Craft of Research the authors (Booth, Colomb and Williams, 2008 p.xiv) comment that they have revised again the chapter on warrants a concept that has been difficult to explain since Toulmin (1958) introduced it.

The Toulmin Model is broken down into the following key elements (Toulmin, 1958, p 94.) which are shown in Figure 2-7.

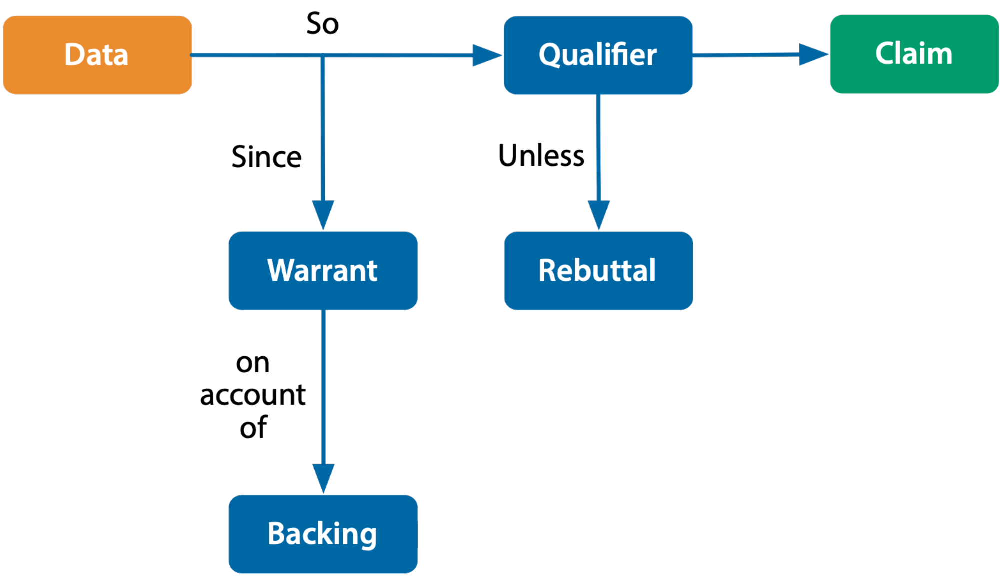


Figure 2-7: Toulmin Model elements with their relationships based on Toulmin (1958, p.94)

A basic evidence-based argument moves from the data (evidence) to a claim. The link between these is often implicit. The aim of the warrants is to make this link explicit which exposes underlying assumptions and principles about how the data is interpreted. The backing justifies the use of the warrant. Disagreements or challenges with the claim are called rebuttals and result in the development of qualifiers. 'Qualification' allows the strength or probability of the claim to be made explicit. Following this model helps ensure that assumptions and principles are made explicit, as well as addressing counter-arguments. This leads to claims that are suitably qualified to guide suitable use and mitigate against inappropriate generalisation beyond what is appropriate.

The Toulmin approach is applicable to a range of types of claims, such as those outlined by Wood (2004, pp.160-16) as follows:

- Claims of Fact: Did it happen? Does it exist?
- Claims of Definition: What is it? How should we define it?
- Claims of Cause: What caused it? What are its effects?
- Claims of Value: Is it good or bad?
- Claims of Policy: What should be done about it?

This research leads to making claims of all these types.

2.4 Research Process

Having discussed the various elements of the approach outlined in Figure 2-1 it is possible to integrate these as an end-to-end process from the problem to a solution (intervention). This is shown in Figure 2-8. The key input and outputs are shown in green. The core elements of the research are highlighted in blue with the wider context of the research shown in orange. In reality, the elements do not fit neatly into such a categorisation but the aim is to aid the 'readability' of the diagram. In addition, the main flow of the research process is highlighted through the use of a thicker line that also highlights the link back to the initial motivation and ultimately the iterative nature of ongoing research in tackling a problem. This conceptualisation is specific to a research process aimed at addressing such a problem, through an intervention, and therefore would differ from research that does not have this objective or did not lead to this as an outcome. This dissertation will follow this process with the specific details of methods used highlighted within the relevant chapters. The terminology of Design Research Method (Blessing and Chakrabarti, 2009) will be used to describe the various phases. The first of these is 'research clarification' which is the subject of the next chapter.

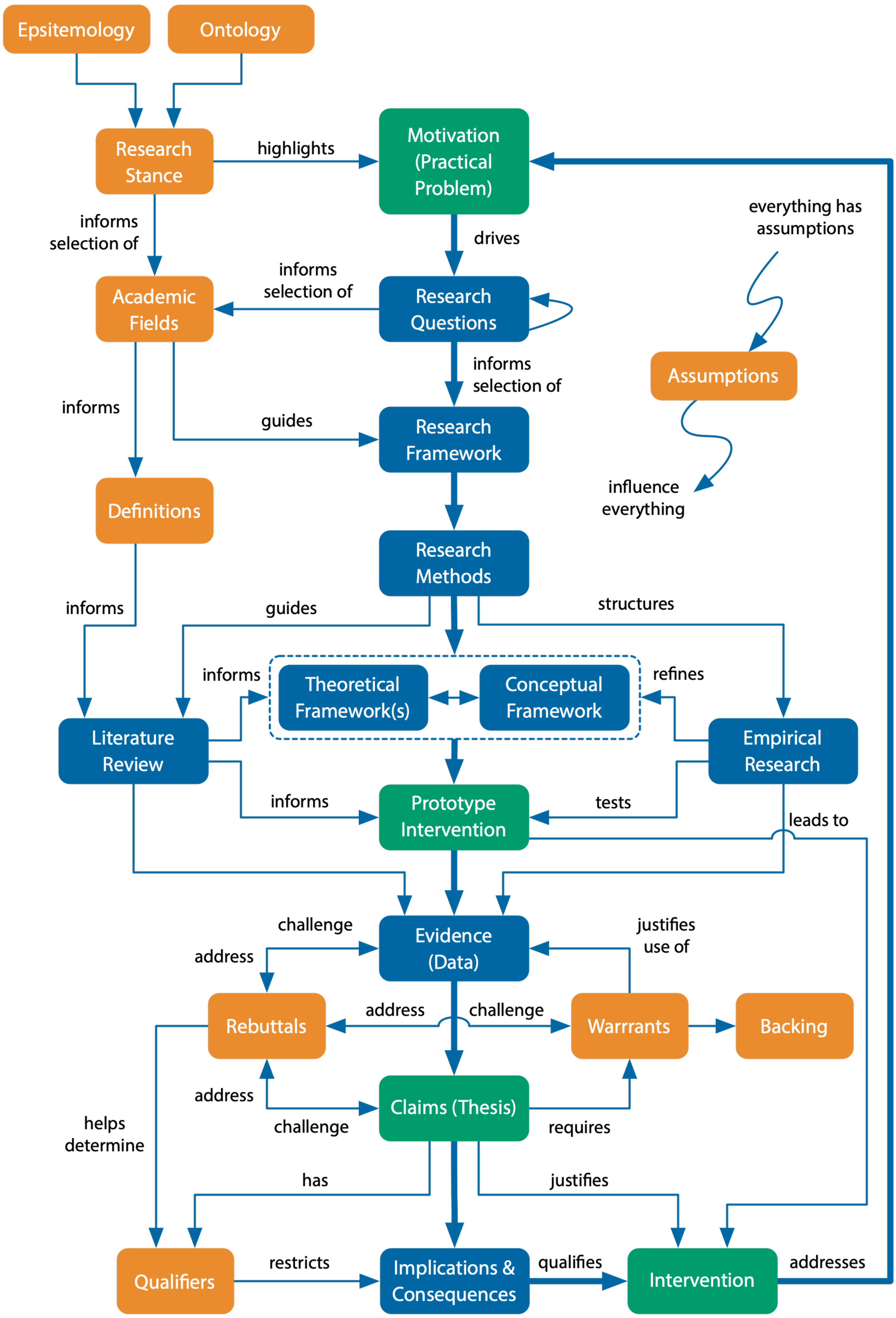
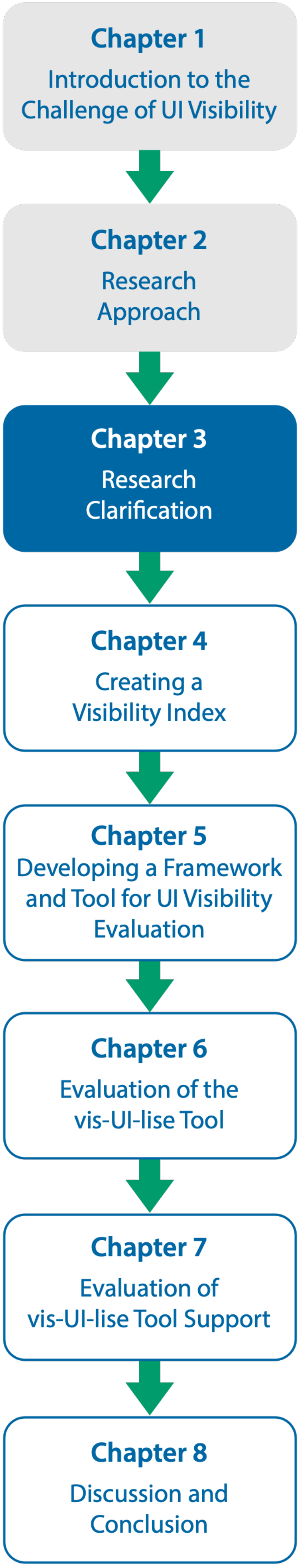


Figure 2-8: The research approach as an end-to-end process (from top to bottom)


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 3

Research Clarification

3.1 Introduction

Having established a worthwhile problem, framed it as a research question (Chapter 1) and chosen the Design Research Method (DRM) as the research framework to address it (Chapter 2), this chapter is what is typically called the literature review. In DRM terms it is called the Research Clarification stage. The chapter aims to establish what is already understood about the problem in the academic literature and to determine if there are any gaps. This then helps frame any subsequent empirical work to address these gaps.

The Research Clarification stage is concerned with finding evidence that supports the *"assumptions"* that the problem outlined can be shaped into a *"realistic and worthwhile research goal"* (Blessing and Chakrabarti, 2009, p.15). This is done primarily through a literature search. In this case, there are two key assumptions. Firstly, that there is such a thing as user interface visibility. Secondly, that it, or a lack of it, causes problems for users. The issue of it being worthwhile was addressed in the first chapter through the socio-technical argument. However, this is only valid if these two assumptions are true.

3.2 Understanding Visibility Issues in Everyday Products

To address these assumptions and understand the phenomenon of user interface visibility **(IRQ 1)** the work followed the approach outlined in Chapter 2 and summarized in Figure 2-6. This consisted of a review of the literature but framed and directed by looking at visibility issues found in everyday products with digital user interfaces. This was done iteratively as described in the previous chapter.

This initial empirical work (as shown in Figure 2-6) consisted of three basic stages as follows:

1. Product selection
2. Exploration of visibility issues (using the selected products)
3. Development of a simple conceptual model (describing the common issues uncovered in the exploration)

The approach to each of these are described in the following sections:

3.2.1 Product Selection

The issue of user interface visibility is potentially of greatest importance in devices used in everyday activities. The notion of 'everyday activities' is potentially nebulous but has been conveniently defined in terms of activities of daily living (ADLs) in work led by Katz. This is comprehensively described by Noelker and Browdie (2014) and expanded by Lawton and Brody (1969) to cover instrumental activities of daily living (IADLs). ADLs cover basic physical activities such as dressing, toileting and feeding, whereas IADLs extend this to cover areas such as communication, shopping and managing the home.

It is primarily the IADLs that form the backdrop for selecting products that could be used in performing or assisting in such activities. This included considering the following product categories of personal computers, transport, domestic appliances and consumer electronics. Also, the diversity of interface types (touch, buttons, pointer-based) and screen sizes from small to large (wristwatch to TV) were considered. Sampling was purposeful concerning the criteria outlined above and convenience with regards to the specific product models. This 'convenience' element was on the grounds of ease of access and cost. This led to the following products being considered which are summarised in Table 3-1 below.

Table 3-1: List of products considered in the initial exploration of user interface visibility issues

Product	Model	Category	Application/ function considered
MacOS	System 6	PC	File management
MacOS	10.9	PC	File management
MacOS	10.9	PC	Address book
iOS on iPad	8	PC	Taking a picture
Toyota Avensis	T4 year 2011	Transport	Heating & ventilation + headlight levelling
Garmin	Forerunner 10	Consumer electronics	Tracking a run
Sharp Microwave	R-92STM	Domestic appliance	Cooking a meal
Virgin Media Tivo Box	SMT-C7100 HW v4	Consumer electronics	Setting a programme to record
Nokia Smartphone	830 WindowsPhone 10	Mobile phone	Taking a selfie and emailing it
Nokia Feature Phone	3310	Mobile phone	Core functions
Apple Remote	1st Gen	Consumer electronics	Core functions
Apple Remote	2nd Gen	Consumer electronics	Core functions
Switch to Scan	Pretorian Technologies 2012	Assistive technology	One button operation of iPad
GoPro	Hero 3	Consumer electronics	Menus
Nikon DSLR	5300	Consumer electronics	ISO setting
Nikon DSLR	D7100	Consumer electronics	ISO setting

Additional factors taken into consideration in the selection were lower and higher feature levels in the same product category and for the personal computer category, an old and new version were considered to look at the changes that had occurred over time. Finally, an assistive technology product was included to provide additional variation to the mainstream products. Two of the products on the list, namely the Switch to Scan and the Nikon D7100™ were reviewed via manuals as the products could not be borrowed but were included as they offered a good contrast with available products. In practice, the exploration became a perspective on using any product that was encountered on a day to day basis by the researcher.

3.2.2 Exploration of Visibility Issues

In examining the products selected, the mindset in the discovery was an open exploration that was abductive in nature, in that it is about 'suggesting' what is going on (Campos, 2009). This orientation has similarities to a grounded approach (Glaser and Strauss, 1967) in terms of continuous reflection of the 'data' but without the more formal elements of coding. It is a process of synthesis (Kolko, 2010) to determine the underlying issues.

The exploration was based on considering specific tasks and the sequence of steps to complete these tasks. Each task step effectively consists of a screen and controls in a specific set of states. To advance to the next step typically required the operation of one or more controls and this became the focus of the exploration. A sample of key screens that were formative in developing the understanding of what was going on are shown as follows.

Example One – Address Book Viewing

The first example is the Contacts application available in MacOS 10.9™ (Apple Inc., 2013) which is shown in Figure 3-1.

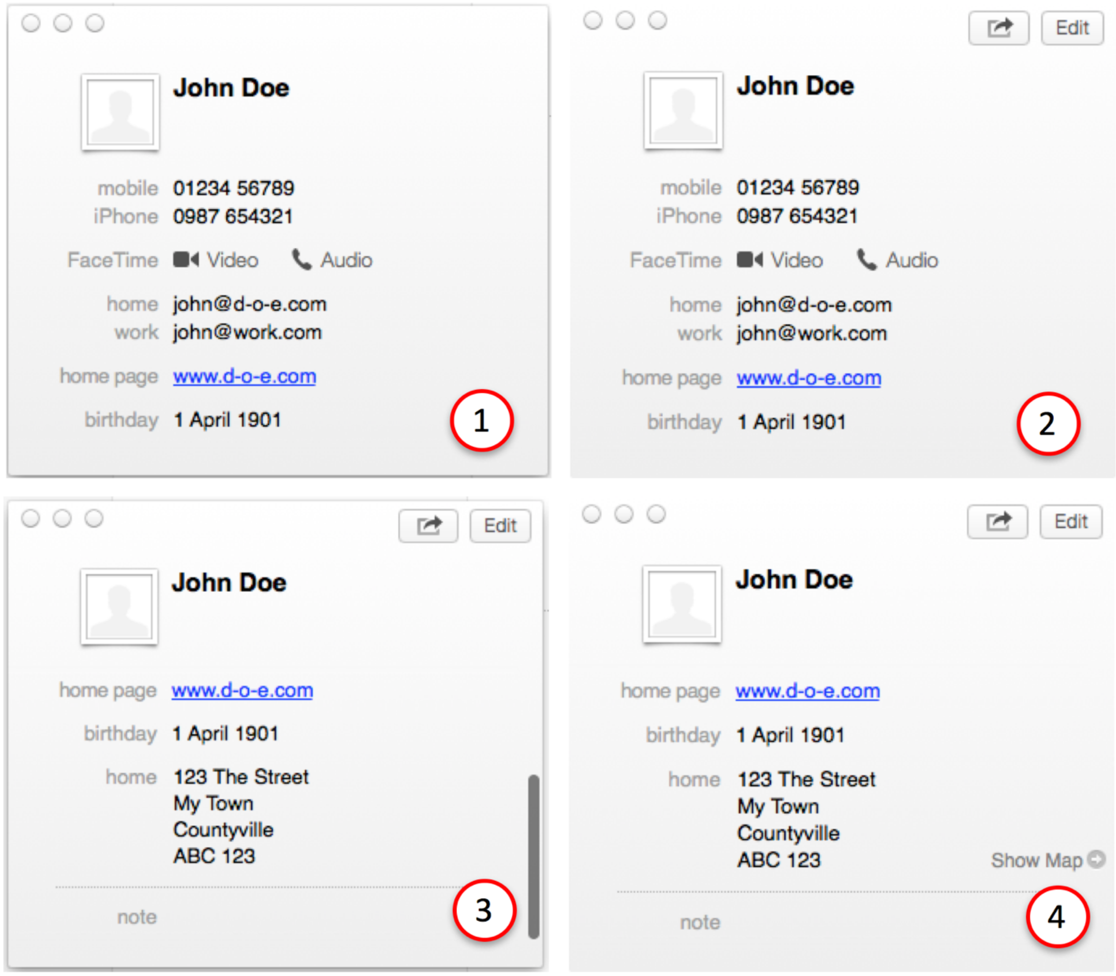


Figure 3-1: MacOS 10.9 Contacts (address book) example from Hosking and Clarkson (2018a)

The numbers in red circles have been added to aid identification of the sequence of steps taken. Image 1 shows an imaginary address book entry that has been opened from the Contacts app. The first 'visibility' test is how to edit or share this entry. The edit and share buttons only appear after the mouse cursor is moved over the window, which reveals the buttons as shown in Image 2. In other words, it has transient visibility. The next 'visibility' test is, whether John Doe's contact record has an address. There is nothing to indicate this (Image 2). If you attempt to scroll, then a scroll bar appears on the right, which is shown in Image 3. Again, visibility is transient, and the presence of an address is only indirectly conveyed by the appearance of the scroll bar once scrolling is attempted. Finally, the ability to open a map is only shown when the cursor is placed over the address and the link itself is displayed to the bottom right of the window (Image 4). The ‘Show Map’ link is also of relatively low contrast. This example highlights that controls can be **‘missing’** from a visibility perspective either on a transient basis, as in this case, or permanently, for which there are numerous examples of gestures on smartphones that have no visible element. Examples of these will be covered in detail in Chapter 4.

Example Two – Picture Editing

The second example, shown in Figure 3-2 below, is editing a photo on an Apple iPad™.

Figure 3-2: Editing a photo on a tablet (iPad Pro 12.9” - iOS 11.4.1)

The screen size on this Tablet is nearly the size of an A4 sheet of paper. The ‘Done’ and ‘Cancel’ buttons are at the very top right and bottom left respectively (these have not been highlighted to enable the reader to have a sense of the challenge). As buttons they have no graphical outline and are just words embedded in the black controls band running along the right edge of the screen. These controls at this screen size are towards the periphery of the user's vision when looking at the picture or picture editing controls (in the centre of the controls band). When the picture is edited the word 'Done' changes to green. This is shown in Figure 3-3 below. These are crops of screenshots of the top right of the screen. This subtle change could easily be **'missed'** particularly as the photographic image itself can change quite dramatically as a result of the edit being performed, and therefore grabbing the user's attention more than the word changing colour.

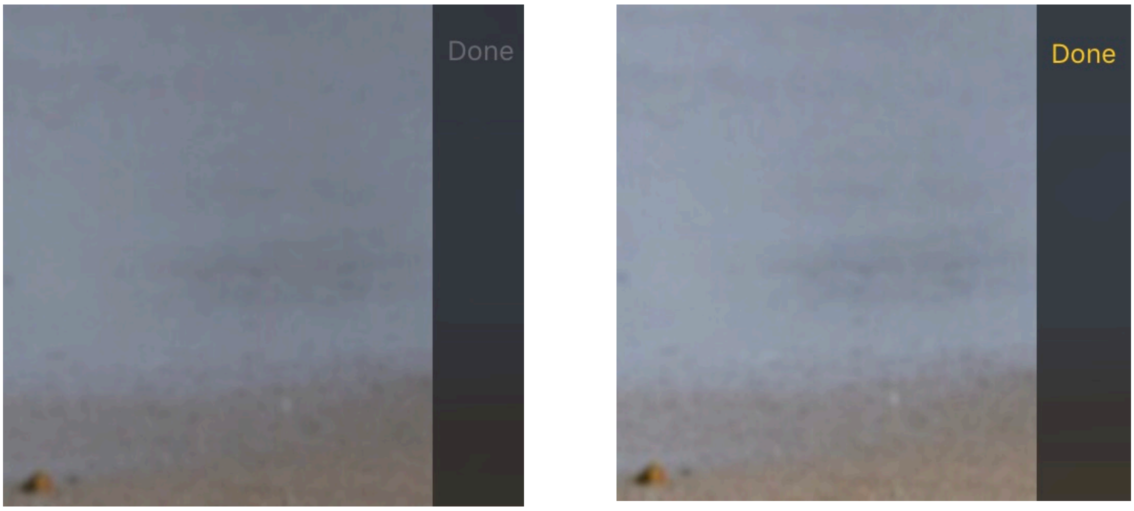


Figure 3-3: Crop of the screenshot showing the change in the colour of the word 'Done' before and after editing has been performed on the image

Example Three – Car Headlight Levelling

The third example is the level control for a car's headlights. Control of the angle of the headlights is provided to compensate for the car carrying heavy loads that causes the car to angle up slightly at the front due to the load on the rear suspension. The control provides 0 to 5 levels in half increments.

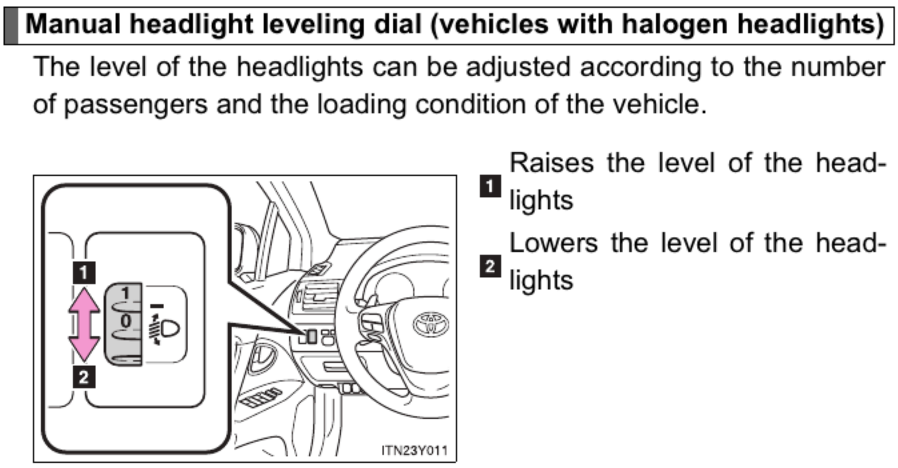
The dial rotates vertically within the dashboard. The question is what level is the dial set to for the example shown in Figure 3-4 below?

Figure 3-4: Headlight levelling set to '0.5' but could be confused as '0'

A cursory glance may lead to the conclusion that the control is set to '0' but it is actually '0.5'. Closer inspection shows that the 'mark' indicating the set value is a horizontal line above the 'light levelling' icon. The user may have an expectation from other controls that the value set is indicated by the centre of the control. This expectation is compounded by the fact that the light levelling symbol is central to the control creating visual clutter with the level line that is above it. Figure 3-5 below shows where it is actually set to '0'. Therefore, the control is seen but potentially **'misunderstood'**. This type of problem is particularly worrying for safety-critical systems such as medical devices. In this case, it may distract the driver, bearing in mind that these photos remove the context of looking down at the lower right part of the dashboard which has a myriad of other controls. Also, the driver may unintentionally not set the level to zero when this would give the best illumination of the road.

Figure 3-5: Headlight levelling set to zero

It appears that this potential confusion may have been shared by the graphic artist for the owner’s manual (Toyota, 2010, p.248). Figure 3-6 below shows a section showing the level adjuster control. The graphic shows that level set at ‘0.5’ but giving the appearance of it being zero.



Manual headlight leveling dial (vehicles with halogen headlights)

The level of the headlights can be adjusted according to the number of passengers and the loading condition of the vehicle.

1 Raises the level of the head-lights

2 Lowers the level of the head-lights

Figure 3-6: Section from the owner’s manual showing the headlight level at an apparent level of ‘0’ but is actually ‘0.5’

Example Four – Car Wiper Delay

A fourth example shown in Figures 3-7 and 3-8 again highlights another example of something that can be 'seen' but potentially **'misunderstood'**. It shows the complex challenge of effectively representing functionality visually and the potential sources of confusion. In this case, it concerns a steering column stalk control that includes a rotating dial (ring) around the axis of the stalk. This is highlighted with a yellow rectangle in Figure 3-7. The control allows the intermittent wiper rate to be changed. Note the different position of the control as seen by the change in location of the word 'AUTO' in the centre of the image. The question is, for Figures 3-7 and 3-8, which setting is a faster rate? Or more specifically what does the increasing thickness in the line between the settings represent? Is it showing an increase in the **time delay**, so the wipers activate less frequently or is it showing an increase in the **frequency** of wiping (shorter time delay) so that they activate more frequently? Two possibilities, that are the opposite of each other.

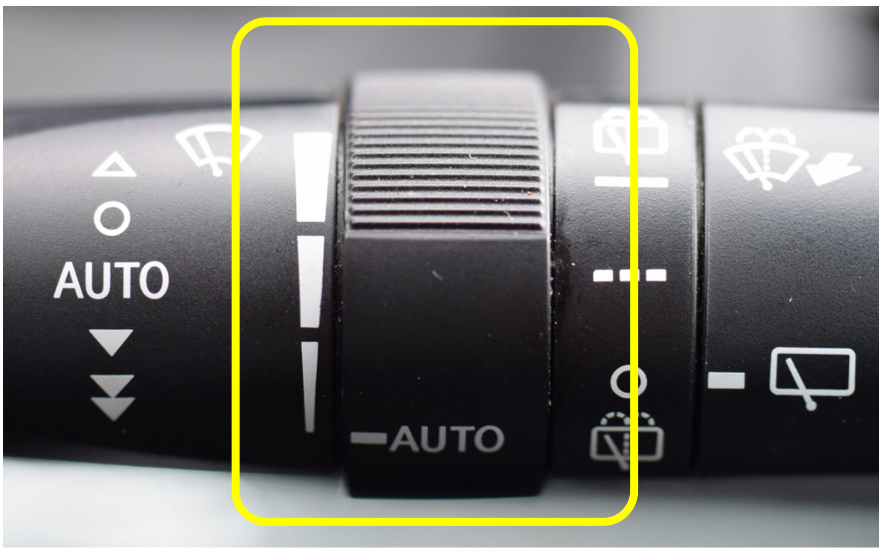
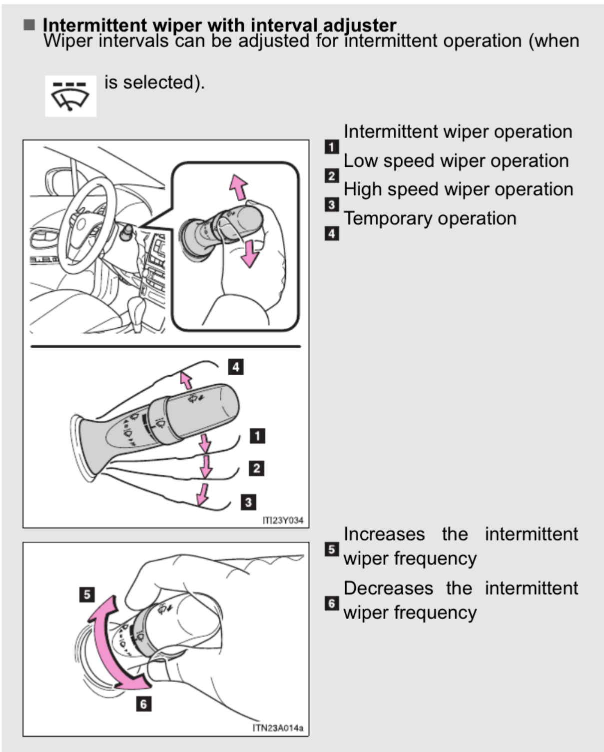


Figure 3-7: Wiper intermittent rate set at its lowest level

Figure 3-8: Wiper rate set at one level above the lowest level

It actually represents an increase in the frequency of wiping (shorter time delay). The manual again potentially causes confusion. Figure 3-9 shows an excerpt from the manual (Toyota, 2010, p.253). The section starts by talking about the 'interval' being adjusted, then the 'speed' of the wiper (non-intermittent settings) and finishes by talking about 'frequency'.


■ **Intermittent wiper with interval adjuster**

Wiper intervals can be adjusted for intermittent operation (when is selected).

1 Intermittent wiper operation
2 Low speed wiper operation
3 High speed wiper operation
4 Temporary operation

5 Increases the intermittent wiper frequency
6 Decreases the intermittent wiper frequency

Figure 3-9: Excerpt from the owner's manual describing wiper interval setting

The potential confusion is compounded by the lack of immediacy of feedback in operation. In other words, the driver has to wait and determine if the rate has increased or not by observing the delay between two or more wipes.

It is also worth noting that this control sits in the context of a stalk that controls other things namely the non-intermittent wiper on/off and speed, the rear wiper operation and front and rear washers.

These four examples are not just esoteric problems from rarely used products. They come from Apple and Toyota both in the top 20 biggest publicly traded companies in the world (Stoller, 2018) and the biggest in their respective sectors. During 2018 Apple stated that they had shipped their 2 billionths device based on their mobile operating system iOS™ (Apple Inc., 2018). These examples of reading an address book, editing a photo, adjusting headlights and wipers are common tasks, yet these examples exhibit significant visibility issues and in the case of the car it is coupled with confusion in the user manual. The impact of such issues may be relatively benign, but the car example moves into an area of functionality that is safety-critical with regards to the driver 'seeing' and 'being seen' (wipers and lights).

3.2.3 Development of a Simple Model

The exploration of visibility issues in the selected products helped drive the development of a model to explain user interface visibility. In practice, it was an iterative process in conjunction with the process of discovery. This modelling concerns getting to the essence of the complex and variable domain of user interfaces into a form that provides a useful description and structure for on-going enquiry. Although the model is a key product of the process, the process itself generates insights and understanding which helps to drive further exploration of the literature.

The first example shows instances where controls are **'missing'** in other words they have no visible components. The second is where a control can be **'missed'** due to the subtlety of its graphical representation and its location on the screen. The third and fourth examples show where a control can be clearly seen but potentially **'misunderstood'**. The examples span from the invisible to the visible. These aspects are brought together in a simple model in Figure 3-10 and the use of the words 'missing', 'missed', and 'misunderstood' lead to it be conveniently called the Three M's Model (Hosking and Clarkson, 2018a).

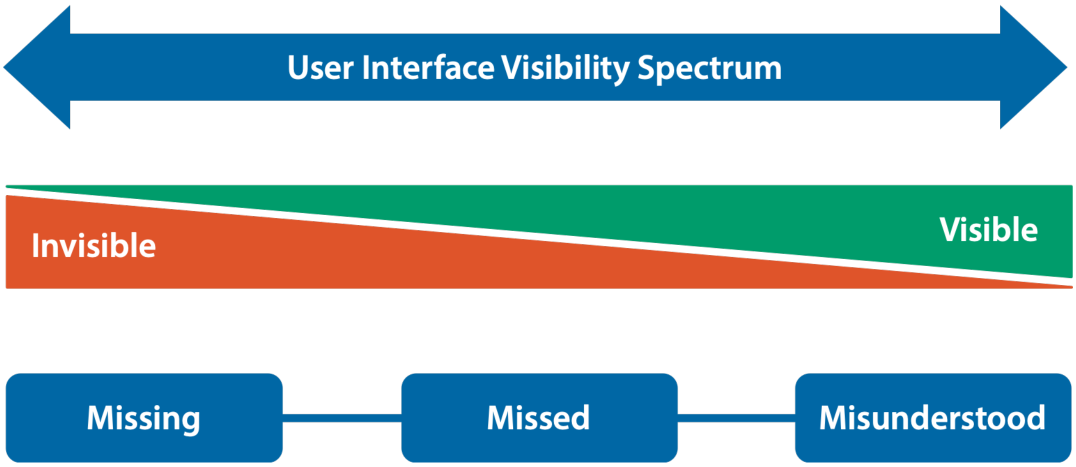


Figure 3-10: The Three M's Model of user interface visibility

The model is in some regards simplistic, for example, 'misunderstanding' can occur across the spectrum. However, a trade-off has been made between 'completeness' of the model and 'comprehension' by both academics and practitioners. The focus is on the model capturing the issues encountered in the examples shown and the many others that were explored during this part of the research. Crucially it challenges the view of visibility that it is solely about whether the user has sufficient visual acuity to 'see' a control. This view is clearly inadequate when a control is invisible ('missing'), but also it does not address the psychological component ('misunderstood'). It could be argued that most of this should be covered separately by a cognitive view of the user interaction, but it rightly raises the question of, where does the process of vision start and end? This issue will be addressed later in the section on vision. This initial product exploration reinforces the view that there is more to UI visibility than 'meets the eye'. Overall this model is a fundamental framing for the literature review and the research as a whole. It begins to answer the Initial Research Questions regarding user interface visibility (IRQ1) and the problems it can cause users (IRQ2).

3.3 Developing the research questions

The Three M's Model helps pose supplementary research questions (SRQs) to help understand the initial research questions (IRQs). This is structured as a hierarchical breakdown starting with the overarching research question (ORQ) which leads to the three initial research questions set out in Chapter 1. These are underpinned by supplementary research questions as follows. With regards to the first initial research question:

> **IRQ 1:** *What is UI visibility?*
>
> This requires an understanding of:
>
> **SRQ 1a:** *What is a user interface?*
> **SRQ 1b:** *What is vision?*

SRQ 1b encompasses the issue of where vision 'starts and ends' mentioned previously. The second initial research question also needs augmenting as follows:

> **IRQ 2:** *What problems is it causing users?*
>
> This requires an understanding of:
>
> **SRQ 2a:** *What problems can it cause?*
> **SRQ 2b:** *Are these problems significant?*

With regard to 'what problems can it cause?' (SRQ 2a) this is addressed in part by the Three M's Model that provides three convenient categories of problems derived from real examples. It is harder to quantify if these problems are significant in real-world use. Although, section 3.2.2 highlights four real examples from major products, of which two of the examples are safety-critical as they relate to driving a car.

There are then a further set of supplementary questions that support both IRQ 1 and 2 that focus on usability as follows:

SRQ 1&2a: *What is usability?*
SRQ 1&2b: *What is a usability problem?*
SRQ 1&2c: *How is usability evaluated?*
SRQ 1&2d: *How well do these approaches address visibility?*

The elaboration of the research questions is summarised in Figure 3-11. This includes blanks for supplementary research questions to the third Initial Research Question, these SRQ's are defined as part of the research clarification process covered later in this chapter.

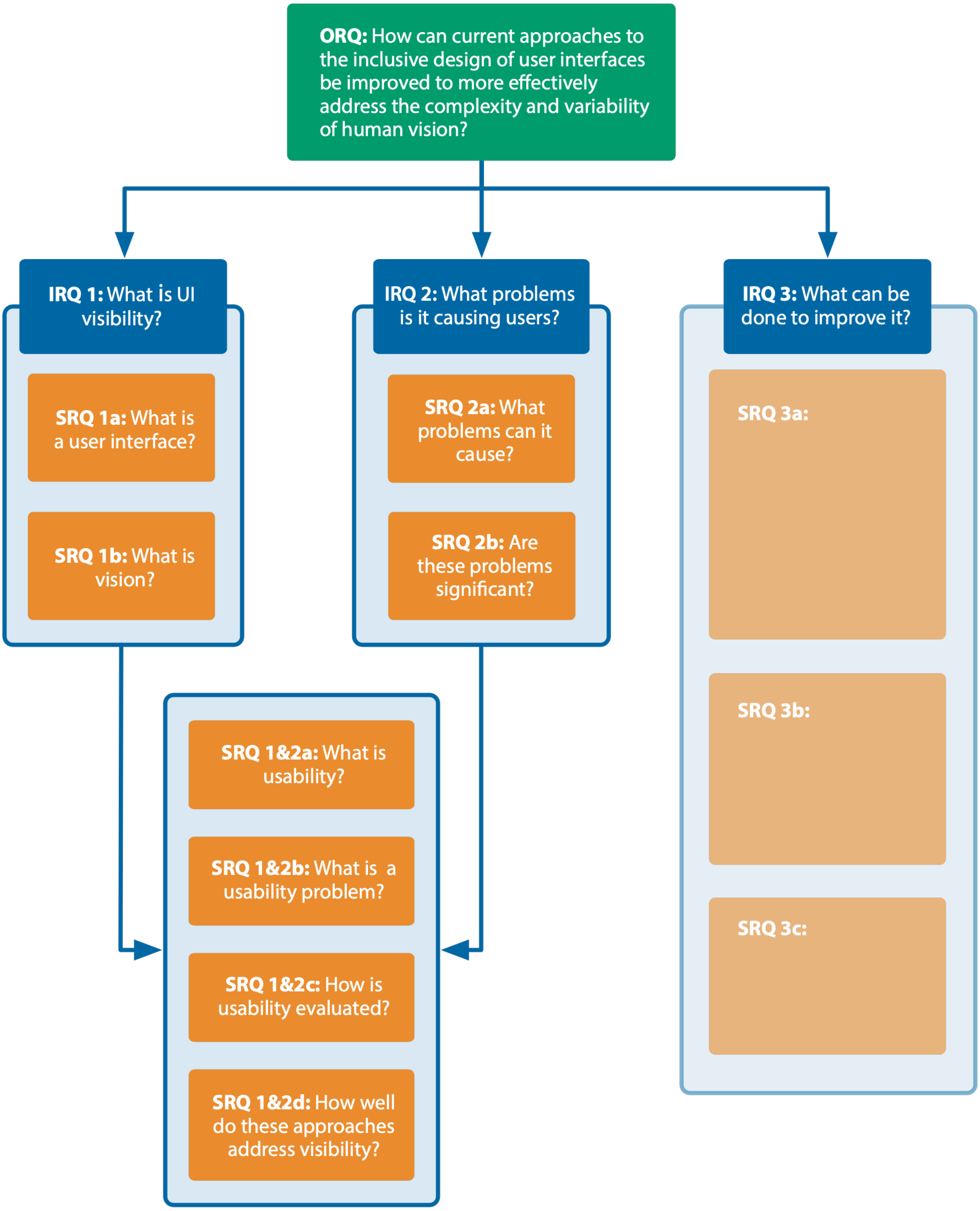


Figure 3-11: Elaboration of the Initial Research Questions 1 and 2

3.4 Reviewing the Literature

3.4.1 Structuring the Different Literature Areas

The Three M's Model with the elaborated research questions help structure what areas of the literature needed to be reviewed. An Areas of Relative Contribution (ARC) diagram (Blessing and Chakrabarti, 2009, pp.63-66) was used to provide a visual structure for the work. This is shown in basic form in Figure 3-12. The primary contribution is positioned in the field of inclusive design. This is based on the view that tackling poor UI visibility could be a key driver for inclusion amongst the elderly and it reflects the research group within which the research was situated. The latter is key in terms of contributing to other related work within the research group and potentially a broader contribution over time.

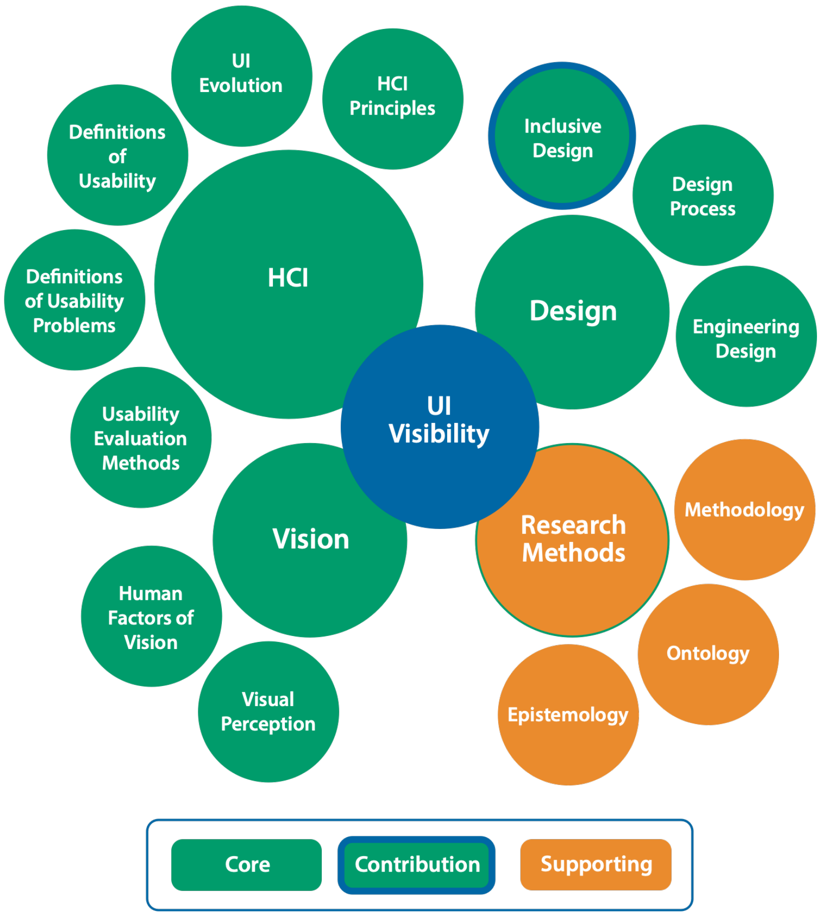


Figure 3-12: ARC diagram showing keys areas of the literature

The ARC diagram was annotated with key or seminal papers in each contributory area of the diagram which is shown in Figure 3-13. This is a heavily 'stylised' version of the original to make it more readable in the format of this thesis. The original was very much a working document for use in 'digital' form to allow viewing of the landscape of the literature. This proved to be a useful visual index to the literature and to see the links across different areas and help structure thinking about the issues. As such the papers on the diagram represent the *"initial pearls"* (Hinde and Spackman, 2014) discussed in the previous chapter concerning the approach to the literature review. They formed the basis for subsequent *"snowballing"* of the literature to get a more comprehensive view of it.

During this process, work shifted from the ARC diagram to a table format, where key points for each paper were written and these summaries were used to form the content of this chapter. It is worth noting a number of things about this initial view. Firstly, there is a significant variation in the number of papers in each area of contribution. This is particularly evident on the left versus right-hand side of the 'stylised view' in Figure 3-13. The areas with more papers arose for a number of reasons, including ones that were more problematic to understand and where they were critical in driving understanding and the direction and structure of the research. Secondly, the process was highly iterative, and the 'snowballing' often resulted in moving significantly further than the 'initial pearl' may have indicated, indeed some of the 'pearls' were not used in the final literature review write-up.

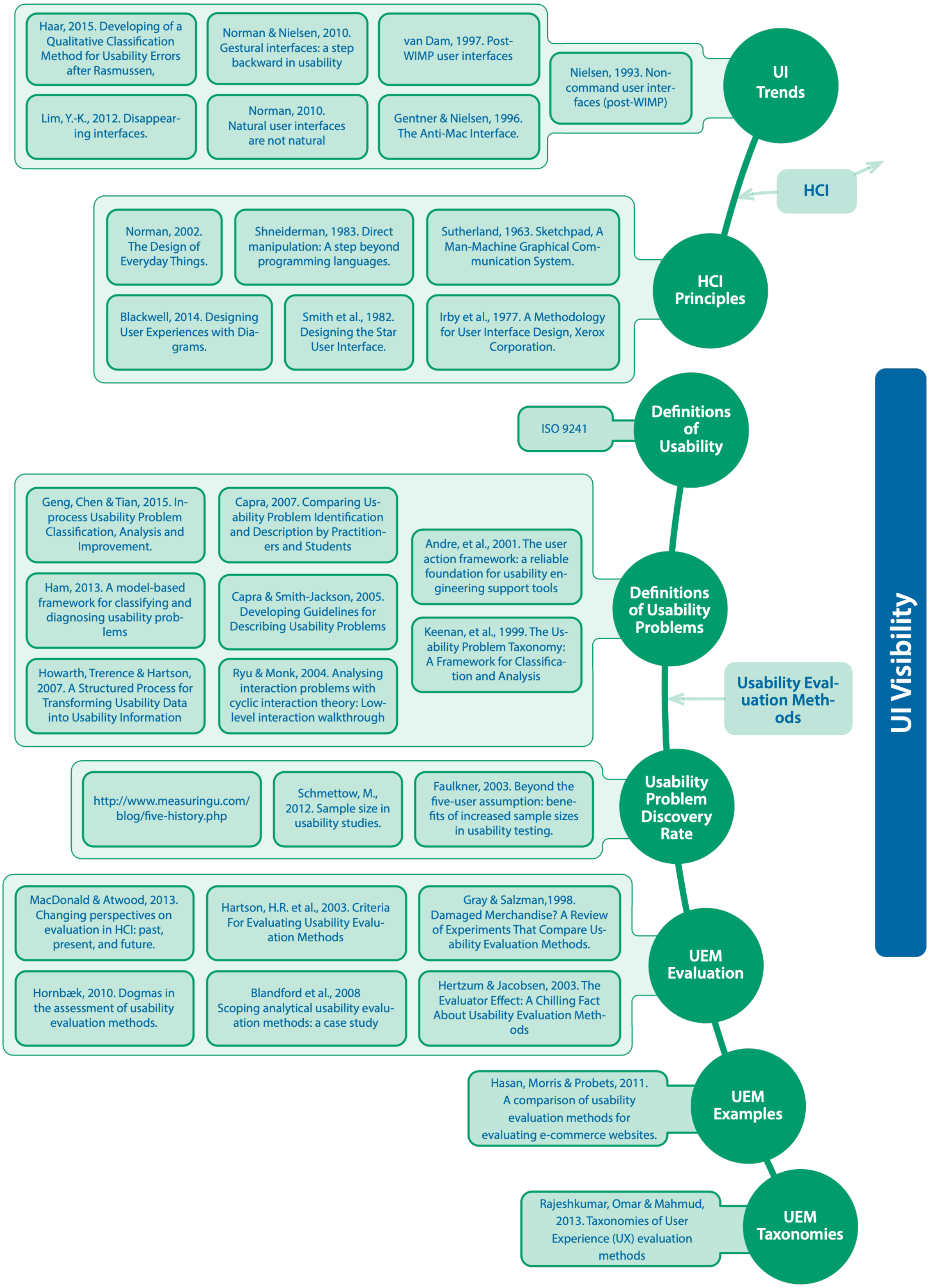


Figure 3-13: ARC diagram annotated with initial papers (left side)

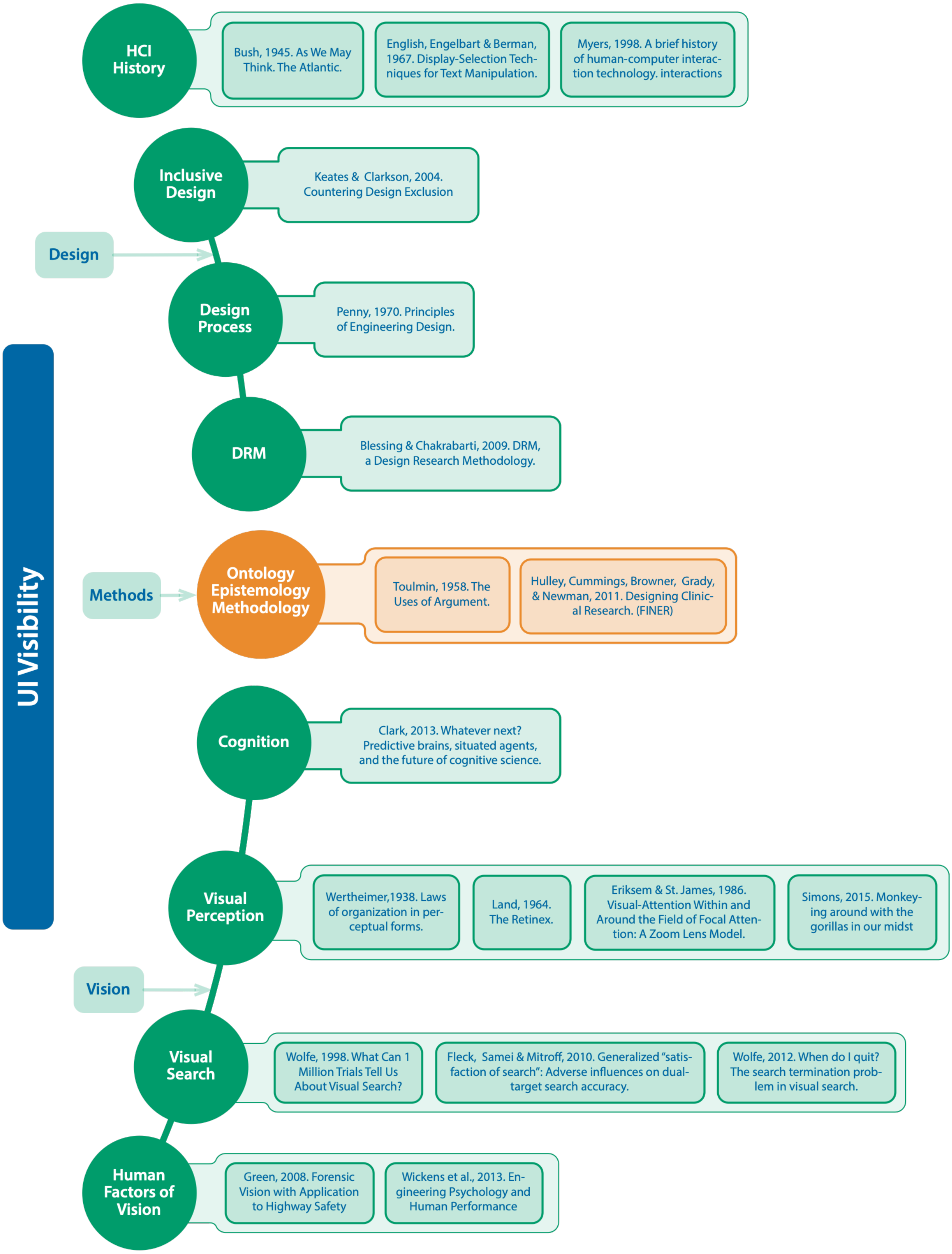


Figure 3-13: ARC diagram annotated with initial papers (right side)

3.4.2 Defining the detailed process

The ARC diagram provides a simple way of identifying and structuring the literature. However, the review itself requires a robust process which was described in detail in Chapter 2 and summarised in Figure 2-6. This is further elaborated in Figure 3-14 to show the iterative interaction between the empirical work already described in this chapter and the exploration of the literature. The mapping of this to the relevant research questions is shown in Figure 3-15. These research questions form the structure and headings for the following analysis of the literature.

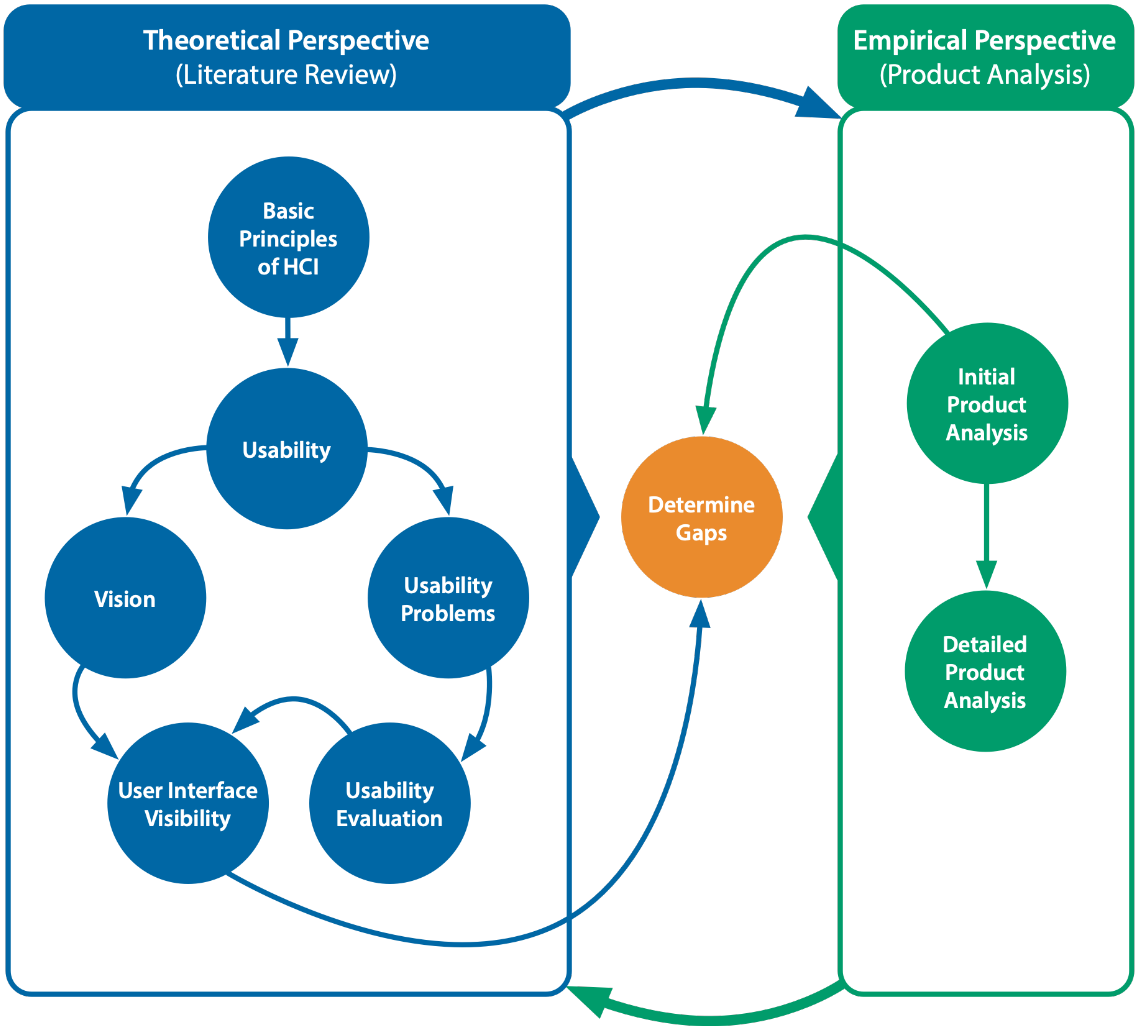


Figure 3-14: The Research Clarification process comprising literature exploration in key areas in conjunction with initial product analysis

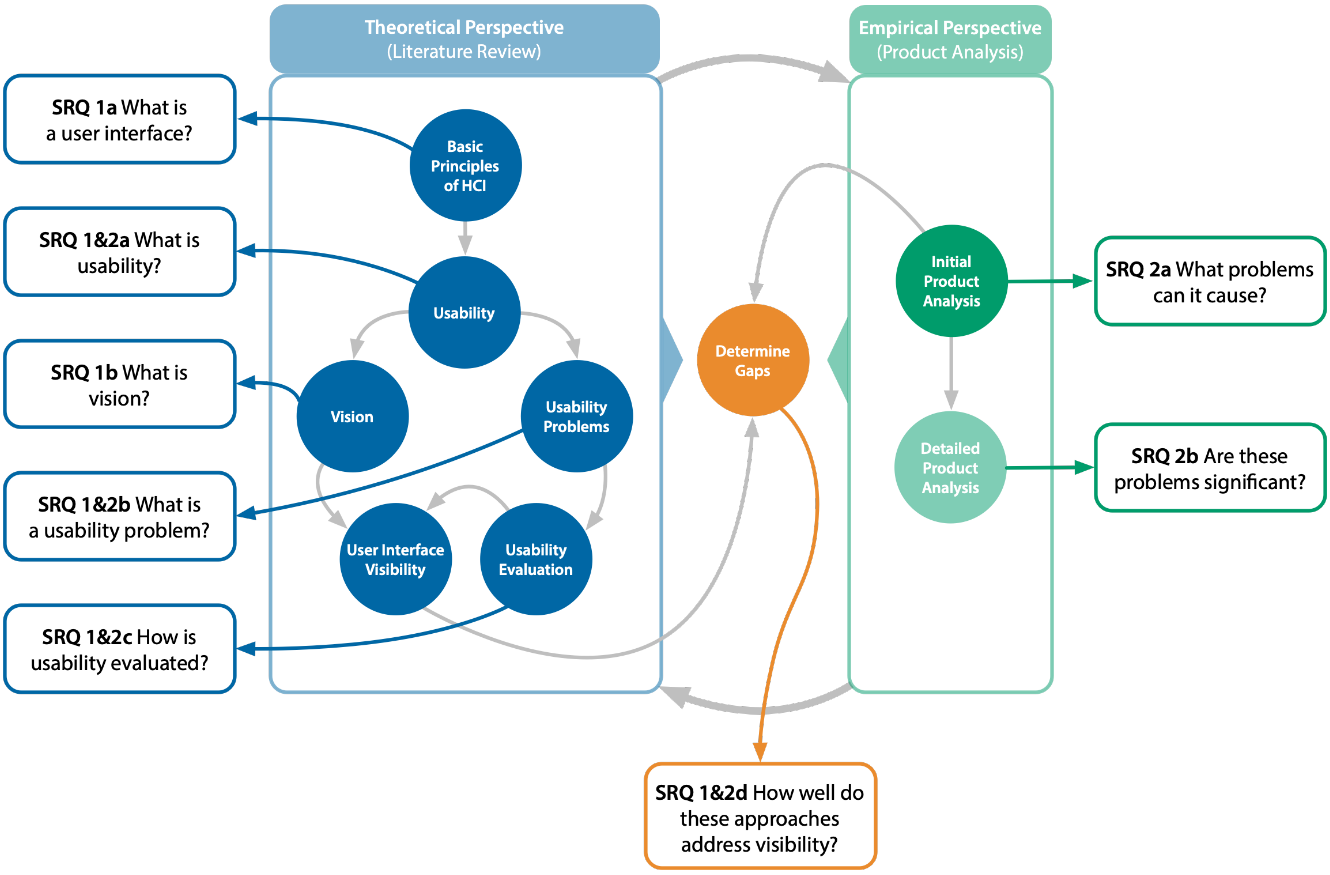


Figure 3-15: Mapping of research questions to the Research Clarification process

3.5 What is Vision? (SRQ1b)

From early on in the research it became apparent that there is more to UI visibility than 'meets the eye'. At the core of this, it is important not only to understand vision but also people's understanding of vision. The latter is important as potentially a poor model of vision might result in usability practitioners not considering fully the needs of users in their designs. This all ties into the Overarching Research Question (ORQ) in terms of current approaches, the goal of improvement and the context of the complexity and variability of human vision.

> **ORQ:** *"How can current approaches to the inclusive design of user interfaces be improved to more effectively address the complexity and variability of human vision?"*

A brief overview of the history of vision science shows that early models of vision were completely the 'wrong way round' in that it was believed it happened by emissions emanating from the eyes as opposed to what we now call 'light' entering them. This started with Alcmaeon of Croton over 2500 years ago proposing that the eye was made up of fire and water and was developed further by Plato (427 to 347 BCE) into what is called the extramission theory of vision, i.e. that vision is due to something emanating from the eye (see Gross, 1999; Debernardi et al., 2010). This was challenged by the likes of Democritus (460-370) who postulated that vision was due to a layer of atoms emanating from the surface of objects, called eidola, entering the eye (see Gross, 1999; Tsoucalas, Karamanou, Kousoulis and Androutsos, 2012). Then Hermholtz proposed the 'sign theory' of perception where the brain makes 'unconscious inferences' to create a coherent view from what is sensed, what we now call a psychophysical view (see Patton, 2016). Current work includes taking a Bayesian probabilistic approach, put simply that vision is the brain's best guess of what is happening using prior knowledge of the world (Adams, Graf and Ernst, 2004).

What is perhaps surprising is that although our understanding has advanced and the model has turned around to it being based on light entering the eye, studies have found that children can develop similar extramission models of vision (Dedes, 2005) and in other studies that such models can persist in adults (Winer et al., 2002). This historical perspective reinforces the view that vision is complex, secondly that it has a strong cognitive dimension, and thirdly it is easy to develop highly inaccurate models of how it works. Understanding what usability practitioners understand about vision could be significant in addressing issues of poor UI visibility.

Even when a person has a reasonably robust 'optical' understanding of vision, there is still the potential belief that 'seeing is believing'. In other words what we see is formed by eyes that act like cameras projecting images into our brain, not considering the vital role of cognitive processing in generating what we 'see'. Green et al. (2008) argue that this 'naïve realism' does not address the fundamental complexity of vision in understanding the visibility of something and they coined the phrase *"believing is seeing"* to challenge this naïve model of vision. Sekuler and Sekuler (2000, p.979) summarise the multifaceted nature of vision well with the statement that *"vision is not a single unitary function, but a collection of separable ones"* and that vision is not just about the eye but depends on contributions from memory and cognition.

Rather than cameras projecting into the brain, a much better metaphor is that of a spotlight which neatly summarises the dynamic and selective nature of vision. It has evolved from the work of James (1890) to the description as such by Posner (1980) and extended to a zooming spotlight by LaBerge (1983). In part, vision is a spotlight due to the narrow field of view afforded by the fovea that gives high acuity, colour vision over an angle of around 5 degrees (Strasburger, Rentschler and Juttner, 2011). The variation in acuity across the visual field is shown in Figure 3-16 and if represented in 3D (horizontal and vertical field of view) it is like a 'hill' which is an analogy first proposed by Traquair (Traquair, 1927; Grzybowski, 2009). Acuity drops off dramatically with increasing eccentricity from the foveal region.

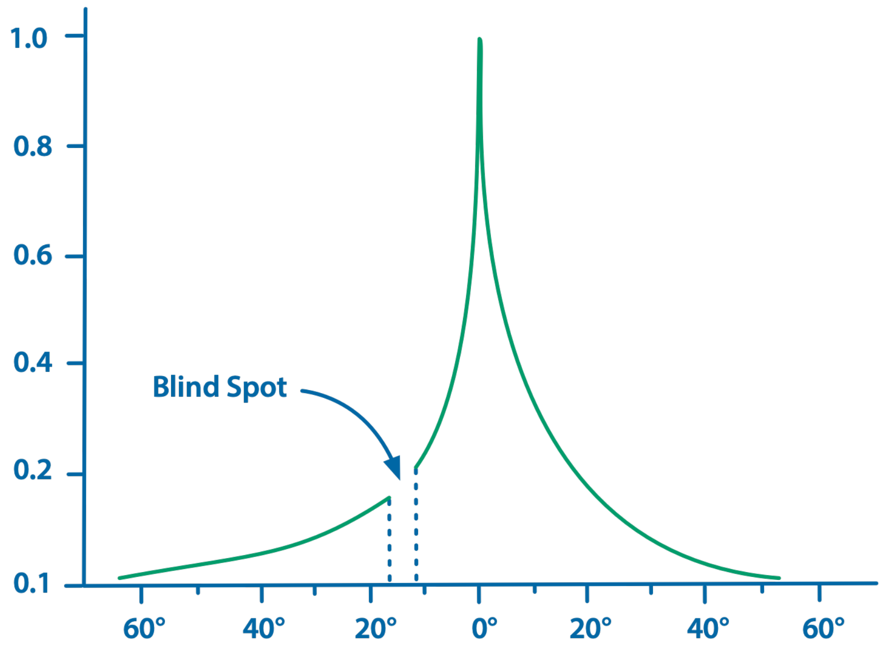


Figure 3-16: Variation in visual acuity across the visual field. Redrawn from (copyright) Vanessa Ezekowitz [CC BY-SA 3.0] and described as Coren's acuity graph by Blanke and Bajaj (2002).

However, other visual attributes vary in different ways with increasing eccentricity due to the heterogeneity of the retina with regards to cone and rod density (Hansen, Pracejus and Gegenfurtner, 2009; Eckstein, 2011). For example, colour vision falls away and beyond around 25-30 degrees has limited practical sensitivity. Other attributes such as motion (fast), luminance, and flashing work well across the whole visual field (Gutwin, Cockburn and Coveney, 2017). Put very simply the spotlight (foveal region) enables high levels of detail to be obtained while the periphery enables the attention of the person to be grabbed so that the spotlight can be put upon other items. The obvious example being objects moving towards (motion) the observer that need to be avoided, such as another person or car. Another way of conceptualizing this is in terms of the different zones that the different parts of the retina afford. Figure 3-17 shows this in terms of the horizontal angle of view. In reality, this is a major simplification but highlights just how narrow the foveal spotlight is. The exact angles of the names of the zones vary across sources (Solso, 1996; Strasburger, Rentschler and Juttner, 2011) but the broad principle is the same.

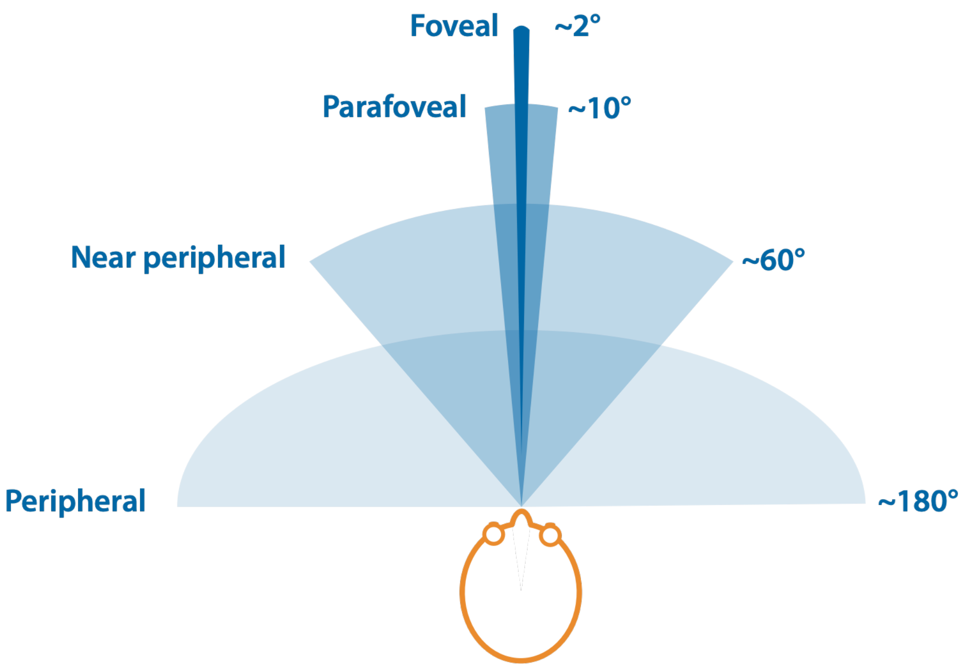


Figure 3-17: Representation of high acuity spotlight – modified from Solso (1996, p.24)

For the observer to make the best use of this valuable, but narrow, high acuity zone the observer can move their head. However, the eye itself conducts several elaborate different types of movement to maximise the use of the foveal region. Glimcher (2003) describes how each eye's six muscles produce five classes of movement: the vesibulo-ocular, optokinetic, saccadic, smooth pursuit and vergence systems. These eye movements are so critical that in experiments that stabilise the image, to effectively eliminate movement, the observer experiences fading of the target within a few seconds until it disappears (Coren and Porac, 1974). Not only are these movements essential and complex but they introduce limitations. For example, with saccades, vision is suppressed during each 'jump' from one fixation point to the next (Bridgeman, Hendry and Stark, 1975; Posner, Snyder and Davidson, 1980), which is why when we look in the mirror we cannot see these movements.

So the spotlight 'dances around' and more recent research has considered the likelihood of multiple spotlights (McMains and Somers, 2004). The nature of eye movements, retinal heterogeneity, attention, cognitive processing and the potential for multiple spotlights means that we cannot simply consider the spotlight in terms of a fixed 5-degree field of view of the fovea. The zooming spotlight, therefore, is potentially a good representation of this variability. It has been argued that all models are wrong (Box, 1976) but some are useful if applied correctly. The challenge for usability practitioners is that the image in our mind is a constant, uniform, 3D, stabilized image i.e. an apparent homogenous performance across the visual field. The reality is heterogeneity with dramatically differing performance with increasing eccentricity from the centre. Therefore, the attentional spotlight and its elaborations are potentially useful models to distinguish from a naïve notion of this apparent uniform representation of the world around us.

Drawing all this together leads to the following answer to the question of 'what is vision?' as follows:

> It is an edge-detecting, dynamic, very slightly delayed, selective, blank-filling, prior-experience-combining, object-inferring, distance, direction and speed-estimating, action-oriented system. It is part of a prolific inference engine, making sense out of an incomplete, noisy, sensory input. Vision is an attentionally-driven, zooming spotlight that outputs a 3D, colour, stabilised, immersive representation of the world that enables effective action within it. [based on Hosking & Clarkson (2018a)]

Having understood the multifaceted nature of the visual system leads to the question of how can we assess how well this system performs? Understanding the limits of the visual system is a potentially useful way of understanding whether something is visible based on it being within them. The limit or threshold is described by Farell and Pelli (1999) as *"the strength of the signal, as controlled by a particular stimulus dimension, that is required to attain a given level of task performance"*. Of key interest with vision is the contrast threshold. This is a psychometric function as it considers both the stimulus and the behavioural response of the observer

(Sukha and Rubin, 2013). Put simply this brings into account the variability of the human in the system that leads to a probability of detection. This leads to a progressive transition from invisible to visible (Runco and Scott, 2014, pp.129-130) rather than an abrupt threshold. Such an approach can be applied to different dimensions of visual performance. Watson et al. (1986) and Watson (2009; 2016) has conceptualised this as the *"window of visibility"*. This has two axes of spatial and temporal resolution. As they behave independently from each other it forms a metaphorical window e.g. square boundary forming the limits. The notion of this window can be extended further (Watson, 2009) to include other key performance parameters shown in Table 3-2. It should be noted that the range of light intensities that the visual system can operate across is vast through adaption, but this process is slow and at a particular point in time it is much more limited.

Table 3-2 Key vision 'window' limits based on Watson (2009) which is based on Watson et al. (1986)

Dimensions	Limits	Unit
Wavelength	380-780	nm
Spatial Resolution	0-60	cycles/degree
Temporal Resolution	0-60	Hz
Field of View	190 Horizontal 125 Vertical	Degrees Degrees
Luminance (unadapted) Luminance (adapted)	$1{:}10^{2}$ $1{:}10^{13}$	Dynamic range min to max

Once the threshold is exceeded (within the window) then performance rises sharply and remains broadly constant beyond this at, what is known as the suprathreshold levels (Legge, Rubin and Luebker, 1987; Runco and Scott, 2014). This constancy of performance within the 'window' leads to the suggestion that if you are in the window then things are visible and therefore from a usability perspective you should be okay. However, although such measures are helpful, and are psychophysical measures they do not fully address the 3 M's and related phenomena such as inattentional blindness (Simons and Chabris, 1999), therefore they only form part of the picture.

Also, it is important to address the fact that this 'window of visibility' will vary from person to person and on a population basis this variability will be driven by the decline in vision due to ageing and other visual abnormalities (Elliott, Whitaker and MacVeigh, 1990; Chauhan, Tompkins, LeBlanc and McCormick, 1993; Pinto et al., 1997; Gillespie-Gallery, Konstantakopoulou, Harlow and Barbur, 2013). Even without abnormalities reduction in visual capability is a 'normal' part of ageing. This is known as presbyopia and is defined by Wolffsohn and Davies (2018) as:

> *"presbyopia occurs when the physiologically normal age-related reduction in the eye's focusing range reaches a point, when optimally corrected for distance vision, that the clarity of vision at near is insufficient to satisfy an individual's requirements"*

This decline occurs throughout life but becomes critical in the 40s and levels out in the 50s (Koretz, Kaufman, Neider and Goeckner, 1989; Wolffsohn and Davies, 2018). On a global basis, this impacts over a billion people and is set to rise as the population ages. Presbyopia's impact on near vision is particularly relevant to most user interfaces that require operation at this distance. It is possible to correct for near vision with glasses, but this requires the user to have such glasses and that they are to hand. It is estimated that in developing countries that unmanaged presbyopia is as high as 50% and as high as 34% in developed countries and that this impacts task performance (Wolffsohn and Davies, 2018). So, not only is visibility critical in user interface use but the critical aspect of near vision performance is severely impacted with age.

This picture of variability becomes more complex when studied in the context of a task. For example, a study of road signs and markings in a driving task (Staplin, Lococo, Sim and Drapcho, 1989) did show a dramatic increase in the average threshold for older drivers of 2 to 2.5 times higher and rising to 20 for the drivers with the poorest vision. However, the report points out that "*cognitive factors play a significant role in driving tasks previously hypothesized to rely principally on sensory capabilities, with implications for the design of traffic control element countermeasures to accommodate the older driver population*". This is consistent with the argument the 'window of visibility' is a useful starting point but not the only factor in understanding visibility, particularly in the context of tasks and the design of systems to support them.

One area of research that supports this broader perspective is that of visual search. A key concept in understanding how we visually search for things is the notion of parallel and serial processing of items. This was proposed by Egeth (1966) and developed further by others such as Treisman and Gelade (1980). At an experimental level testing centres on what you are looking for, the 'target' and things that hinder this the 'distractors' (Wolfe and Horowitz, 2004). The performance in the search task can be measured in terms of the reaction time (RT) required to indicate whether the target is present or absent. For serial task processing, i.e. where a person has to consider each item in the set individually, the gradient increases with set size. See Figure 3-18 which shows the slopes for two different set sizes where the red line represents a set of half the size with the consequential reduction in the gradient of the slope. Each item in a serial search takes around 20-30 ms where the target is present and 40-60 ms where it is absent (Wolfe, 1998). This figure increases further if the observer must fixate on each item and then the time can increase to 150-300 ms per item (Wolfe and Horowitz, 2017). The gradient of the slope indicates the efficiency of the search.

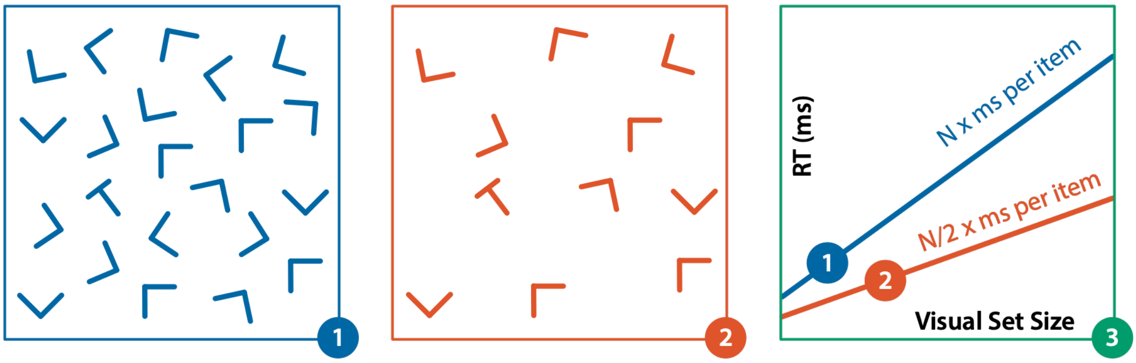


Figure 3-18: A target 'T' is presented amongst a different number of distractors as shown in boxes 1 and 2. The efficiency of the search can be shown by the gradient of the slope of reaction time (RT) with regard to the visual set size (N) as shown in Box 3. The slope for Box 2 is half that of Box 1 because attention is limited to half the number of items [simplified version of Wolfe and Horowitz (2017)].

This becomes very efficient when there is a simple dominant feature such as colour. This is shown by the red '5' in Figure 3-19. Here the search becomes parallel and the gradient is typically flat i.e. independent of set size (Wolfe and Horowitz, 2017). This differs greatly from having to search for the number '2' within the set, which requires a serial search and therefore is potentially much slower.

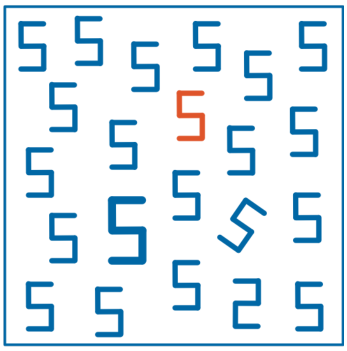


Figure 3-19: Parallel (red, angled and enlarged 5's - easy) and serial (a number '2', which is bottom right - difficult) examples of visual search redrawn from Wolfe and Horowitz (2004)

Figure 3-19 also shows other attributes that can be used to guide attention namely size with the enlarged '5' and orientation with the angled 5. Figure 3-20 shows the use of orientation too, but also the importance of 'localised difference' of the neighbouring distractors. In this case, the central vertical line 'pops-out' and is highly salient. However, closer inspection or serial processing shows that there are in fact five vertical lines in total (look towards the top left corner of the image for the other vertical lines).

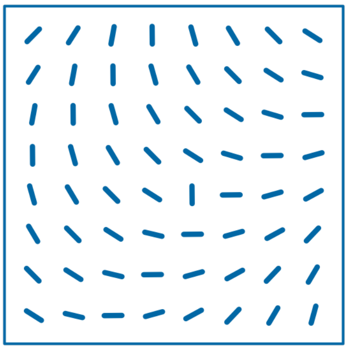


Figure 3-20: An image showing the importance of the difference between the target (vertical line) and neighbouring distractors. Note there are 5 vertical lines in total. Redrawn from Wolfe (2007, p.104).

Wolfe (1998) pointed out from a review of 1 million visual search tests that there is not a simple division between parallel and serial processing, but it is, in fact, a continuum. He states that *"Your favorite theory of visual search is wrong. So is mine"*, but he was optimistic that models can be developed that are better at predicting performance. Indeed, he and his team, has gone on to develop an increasingly more sophisticated model of guided search which is now up to version 5.0 at the time of writing (Wolfe, Cain, Ehinger and Drew, 2015).

This work is been helpfully translated into a pragmatic description of five factors that guide visual search (Wolfe and Horowitz, 2017). This is at a level that is potentially useful for usability practitioners. These are outlined as follows:

1. Bottom-up guidance by stimulus salience

Attention is drawn to items that differ from the ones that surround them. This is shown by the items that 'pop-out' in Figures 3-19 and 3-20. The two fundamental principles of bottom-up guidance are firstly, that salience increases with increasing difference between the target and the distractor, which is known as target-distractor heterogeneity. Secondly, salience improves as the differences between distractors reduce, which is known as distractor-distractor homogeneity.

2. Top-down feature guidance

The bottom-up view is helpful at a basic level of understanding salience, but in the real-world other factors come into play particularly when the observer has a top-down goal. The top-down approach can make use of multiple attributes in guiding the search e.g. colour, size and orientation where efficiency is dependent on if these features are shared by the target and the distractors. Wolfe and Horowitz have an extensive list of attributes categorised from ‘undoubted’, ‘probable’, ‘possible’ and ‘doubtful’, as well as ‘probably not’. The ‘undoubted’ are colour, motion, orientation and size. Interestingly the ‘probably not’ includes ‘colour change’ and a ‘person’s name’. Note that the second example from the empirical work (shown in both Figures 3-2 and 3-3) is an example of a ‘colour change’ that could easily be missed.

3. Guidance by scene properties

Away from lab-based experiments, most visual search takes place in structured scenes. If you are looking for a person then you will do so where they can realistically be e.g. on the ground and not in the sky. A bird, on the other hand, can be in both. If you are looking for a door handle you will do so on the door and around the middle of it. If the door has a release button you can expect it to be near the door also, when it is not you can see why problems start to arise.

4. Modulation of search by prior history

It is not unsurprising that the prior history of the observer modulates the guidance of their attention, i.e. where they will look. Perhaps less obvious is that these effects start from exposures as little as 100 ms. Not only does short priming work, but durations of 200 ms have been shown to last as long as a week.

5. Modulation of search by the value of items

Experiments have shown that value strongly modulates guidance. For example, if an observer has been rewarded more highly for red items than green ones, this will have an impact on subsequent tasks even if colour is irrelevant to the task. This learning can persist over a long period, with value-driven effects being seen 6 months after they were acquired.

An alternative approach to unpacking the complexity of 'targets' and 'distractors' is the notion of visual clutter and how it can be quantified. Rosenholtz et al. (2005) define clutter as, "*the state in which excess items, or their representation or organization, lead to a degradation of performance at some task".* They and others have proposed sophisticated measures for the assessment of clutter (van den Berg, Cornelissen and Roerdink, 2009), but evaluation of five different algorithms against a multi-category visual complexity image dataset showed mostly poor performance compared to the baseline scores (Saraee, Jalal and Betke, 2018). Pankok and Kaber (2018) have developed a clutter model based on a broader range of contextual factors to try and improve the correlation with task performance. Other task-specific measures include those for maps, which highlights the need to address highly context-specific issues to get meaningful correlation with task performance (Stigmar and Harrie, 2013). Such approaches are useful but again only form part of the picture as they only cover performance issues related to clutter and suffer from sensitivity to the context of the task.

In summary, this overview of vision demonstrates several important things. Firstly, it is a highly complex, multifaceted phenomenon and secondly, its performance is highly context dependent. This latter aspect is the key. Simply measuring the legibility of the 'T' in Figure 3-18, the red '5' in Figure 3-19 or the vertical lines in Figure 3-20 will not account for the context of the distractors that surround them, let alone the even greater complexity of the real world.

To thoroughly assess the visibility of a design we need to consider moving beyond a simple threshold model to a context-sensitive, attentionally driven, dynamic spotlight-based view. For Green et al. (2008) such models are critical for the assessment of road traffic accidents where a simple picture of the scene of an accident is insufficient to determine what a person saw. The literature combined with the initial product evaluation points to a similar challenge being required in the design community concerning user interface design.

3.6 What is a User Interface? (SRQ1a)

Having established the nature of vision we need to consider it more fully in the context of user interfaces and to do this we need to understand what a user interface is.

3.6.1 An historical perspective of user interfaces and the role of visibility

A useful perspective on this is to see how user interfaces have evolved and the role of visibility in that evolution. A simple starting point for this is the premise that humans interact or communicate with computers in a similar way to human-to-human communication, as shown by the simple model in Figure 3-21.

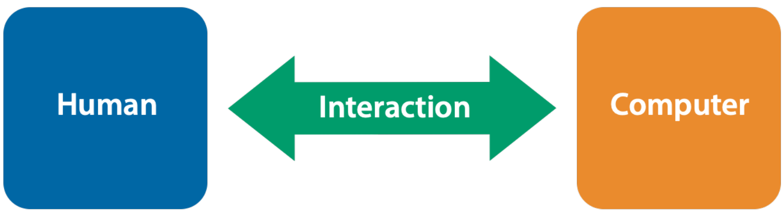


Figure 3-21: Basic model of human-computer interaction

Licklider (1960) in his seminal paper Man-Computer Symbiosis recognised the nature of this interaction and the fundamental shift from purely mechanical devices being simple extensions of human capability, for which he uses North's (1954) description of the 'mechanically extended man'. His thinking went even further to envisage the trajectory towards artificial intelligence all in an era when interacting with computers was done by punching holes in paper cards.

However, such vision was predated by Vanevar Bush (1945) who in his essay in the Atlantic Magazine describe a hypothetical machine called the 'Memex', which was a mechanical representation of the modern World Wide Web. Such vision began to be realised by the likes of Sutherland (1963) with the Sketchpad which can be considered as the first graphical user interface (Myers, 1998) and the point at which visibility becomes a critical component of the interaction. Sketchpad's interaction was driven by a light pen on the screen. Englebart led the team that replaced this input device through the invention of the mouse (English, Engelbart and Berman, 1967; Myers, 1998) and went on in 1968 to demonstrate a well-featured system that has become known as the 'mother of all demos' (Doug Engelbart Institute, 2018). By 1972 Kay had drawn up a detailed vision for a tablet device called the DynaBook (Kay, 1972). By this point, the Graphical User Interface (GUI) was very much born with visibility implicit in the use of the word 'graphical'.

This leads to an important distinction in the approach to the underlying nature of the interaction between humans and computers. This is well expressed by Hutchins, Hollan and Norman (1985) who describe two fundamental metaphors for the interaction. The first is a *"conversation metaphor"* where the interaction is mediated through a language that forms a 'conversation' between the human and the computer about some assumed 'world'. This is typically seen in command-line interfaces. The second is the *"model-world metaphor"*. Here the interface itself is a 'world' that the user can interact with and one which changes its state in response to user actions on it. This leads to a directness of engagement that underpins a GUI and is the dominant form in today's non-technical user interfaces.

Xerox Parc (Irby et al., 1977) played a key role in the development of the GUI including the development of the Star Interface (Smith et al., 1982). The Star Interface was the pre-cursor to the first commercial system the Xerox Star in 1981, which inspired the Apple Lisa in 1982, and the Apple Macintosh in 1984 (Myers, 1998). This trail leads directly to the MacOS and indirectly to many of the other products that were investigated earlier (see Table 3-1).

Smith et al. (1982, p.248) contrast the desirable characteristics of user interfaces with undesirable ones.

Table 3-3: Desirable characteristics of user interface concepts contrasted to undesirable ones. Based on Smith (1982, p.248)

Desirable Characteristics (Easy)	Undesirable Characteristics (Hard)
Concrete	Abstract
Visible	Invisible
Copying	Creating
Choosing	Filling in
Recognizing	Generating
Editing	Programming
Interactive	Batch

Of note is the contrast between 'visible' and 'invisible' and how important visibility was deemed to be in this early work.

This is further emphasized by Smith et al. (1982, p.248) when they describe the main goals that they were pursuing in the design of the Star user interface as shown in Table 3-4 below:

Table 3-4: User interface goals for the Star Interface, based on Smith et al. (1982, p.248)

User Interface Design Goals
familiar user's conceptual model
seeing and pointing versus remembering and typing
what you see is what you get
universal commands
consistency
simplicity
modeless interaction
user tailorability

Goals two and three again emphasize the role of visibility and the first one, to a degree, the psychological aspects of vision. Shneiderman (1983a) provides an elaboration on Goal 3 of *"what you see is what you get"*. He states his view that the success of early graphical user interfaces is due to the "*visibility of the object of interest"* with rapid, reversible, incremental actions controlled by the direct manipulation of the object as opposed to a complex command language. Norman (2002, p.13) is clear on this as well and in his discussion of the fundamentals of interaction he highlights the two key areas of firstly, having a *"good conceptual model"* and secondly, to *"make things visible"*. Norman (2010) also said, *"the important design rule of a GUI is visibility"*.

GUI's have continued to develop and we live in, what has been described as, a post-WIMP [Windows, Icons, Menu, Pointer] era (Nielsen, 1993a; Gentner and Nielsen, 1996; van Dam, 1997; 2001). This is in part due to the growth of touch-based devices. The 2 billion Apple devices mentioned earlier (Apple Inc., 2018) can be categorized as non-WIMP products. However, concerns have been raised about the usability of post-WIMP devices. Norman and Nielsen (2010) describe the situation as a *"usability crisis"*. This alleged "*crisis*" is exacerbated by products that are suffering from a proliferation of features. For example, in Microsoft Word™, the number of commands in Word 1.0™ was about 100 but by Word 2003™ it had exceeded 1500. When Microsoft asked users what they wanted in the next version of Office, 9 out of 10 asked for features they already had in their current version (Caposella, 2005). With such an abundance of features, this could well be because of 3 M's type problems.

Drawing this evolution together Hosking and Clarkson (2017) devised a simple evolutionary model of user interface styles which builds on Smith et al. (1982) goal of *"seeing and pointing versus remembering and typing"* for early GUI's. This is shown in Table 3-5 below:

Table 3-5: Evolution of interface styles based on Hosking and Clarkson (2017)

Interface Style	Description
Command Line	Remember and type
WIMP	See and point
Touch interface	Remember and swipe
Gestural	Remember and wave

This perspective reinforces the concerns about post-WIMP interfaces and the potential loss of visibility. Essentially it can be argued that there is a partial return to the disadvantages of command-line interfaces, rather than having to remember text commands the user has to remember gestures, for example 'pinch to zoom' on a touch-based device.

Overall, this historical perspective shows several key things:

1. Visibility is a key thread throughout the evolution of user interfaces
2. There was an early emphasis on the importance of visibility
3. The evolution to a post-WIMP era has led to real concerns that visibility is diminishing, through the history of user interfaces there appears to be a rise then fall in visibility.

Despite the role of visibility, one of the striking things in the literature is that visibility is emphasised but never defined explicitly. This further reinforces the validity in question IRQ1 ‘What is UI visibility?’ and one which will be returned to later after considering related issues that help answer it.

3.6.2 Models of human-computer interaction

Having established at a basic level that human-computer interaction is a two-way interaction between a person and a computer (Figure 3-21) and that visibility plays a key role in the interaction, it is appropriate to look more deeply at the nature of this interaction. This has been done through various theoretical models of the interaction. Rogers (2004) provides a useful categorisation of types of theories based on the work of Shneidermann (Shneiderman and Bederson, 2003, pp.349-351) these are summarised below.

1. Descriptive - what it is
2. Explanatory - how it works
3. Predictive - how well it might work
4. Prescriptive - how to make it work well
5. Generative - create something new or a new understanding

The following models shown in Table 3-6 were reviewed and categorized in terms of the different types listed above. In addition, a view is given on the main focus i.e. cognitive, vision or action.

Table 3-6: Interaction models reviewed which is an extended version of Hosking and Clarkson (2018a)

Model	Description	Focus	Primary Type
Perceptual Cycle (Neisser, 1976)	A generic, seminal cognitive model of how humans interact with the world	Cognitive	2
Model Human Processor (MHP) (Card, Moran and Newell, 1983b)	Simplified cognitive architecture	Cognitive	2
GOMS Family of Models (Card, Moran and Newell, 1983a; John and Kieras, 1996)	Simplified cognitive architecture for predicting performance related to MHP	Cognitive	3
7 Stages of Action (Norman, 1986)	Model of action that includes the gulfs of execution and evaluation	Action	2
User Action Framework (Andre, Hartson, Belz and McCreary, 2001)	A simplified form of Norman's (1986) 7 stages of action	Action	2
Task-Artifact Cycle (Carroll, Kellogg and Rosson, 1991, p.80)	A simple model showing the relationship between the task and the device	Action	2
Object-Action Interface Model (Shneiderman and Plaisant, 2005, pp.95-101)	Relates the task to the interface in the context of direct manipulation	Action	2
Cognitive Dimensions (Green, 1989)	A set of heuristics for visual design	Cognitive	4
Context of Use (Bevan, 1995)	A high-level model of the context of use to aid the structuring of measures of quality of use	Action	1
ACT-R (Anderson and Lebiere, 1998)	Sophisticated cognitive architecture	Cognitive	3
Three-stage Model of Vision (Ware, 2003)	A simplified three-stage model of vision	Vision	2
Capability-Demand Model of Interaction (Persad, Langdon and Clarkson, 2007; Waller, Langdon and Clarkson, 2009)	Describes how exclusion can occur when the demands of a product exceed a user's capabilities	Action	2
Physics, Physiology, Psychology Framework (Green, 2008)	A broad framework for vision including perceptual exploration	Vision	2
Human Information Processing Stages (Wickens, Hollands, Banbury and Parasuraman, 2013)	Model showing process stages relevant to task execution	Cognitive	2
PCA [Perception-Cognition-Action] Model (CENELEC, 2015, p.24)	Model adopted by standards bodies for medical device development and evaluation	Action	2

These models where influential not only in understanding UI visibility but also the development of the intervention in this work and we will return to some of them later. At this stage is helpful to consider two of these that help highlight key conceptual areas. These are the Model Human Processor (Card, Moran and Newell, 1983b) and the three-stage model of vision (Ware, 2003).

Model Human Processor

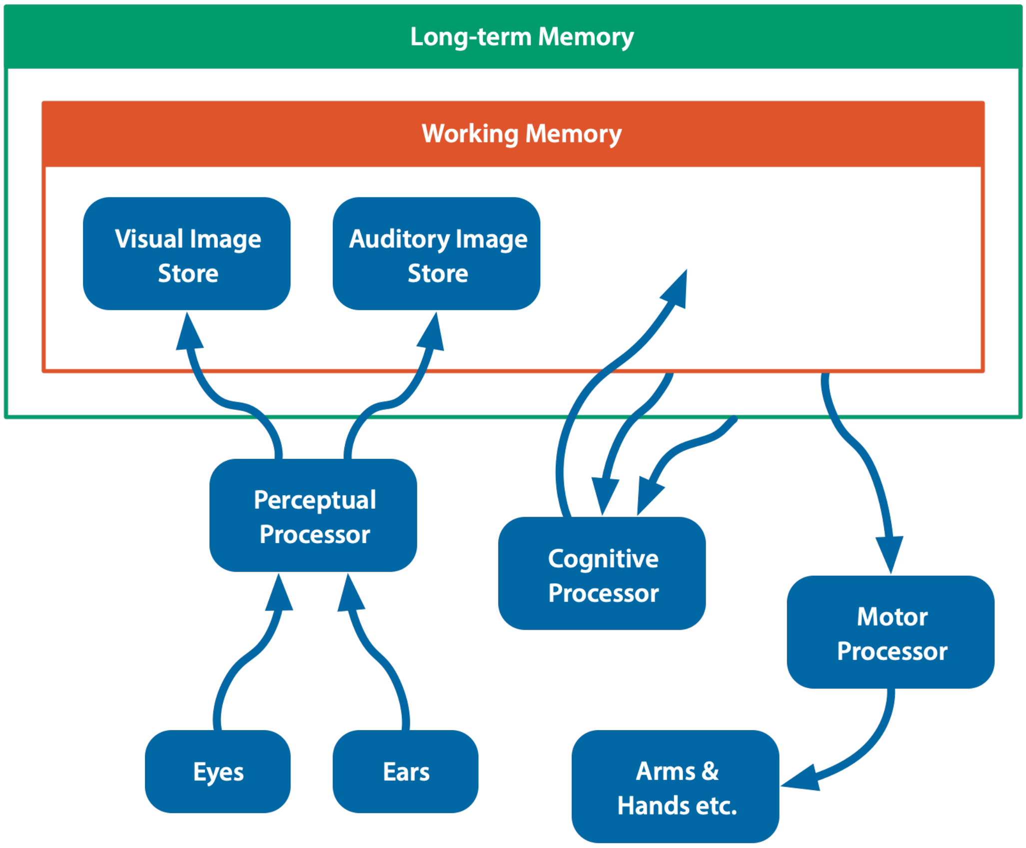


Figure 3-22: A simplified representation of Card, Moran and Newell's (1983b) Model Human Processor

The Model Human Processor (MHP) unpacks the basic notion of human-computer interaction shown in Figure 3-22 and integrates the multiple components (memory and cognition) of vision discussed earlier (Sekuler and Sekuler, 2000, p.879). It also makes explicit the notion of perception within the system. This model is complemented by Ware's (Ware, 2003;

2004, p.2022) three-stage model of vision which extends the vision-specific elements of the MHP.

Three Stage Model of Vision

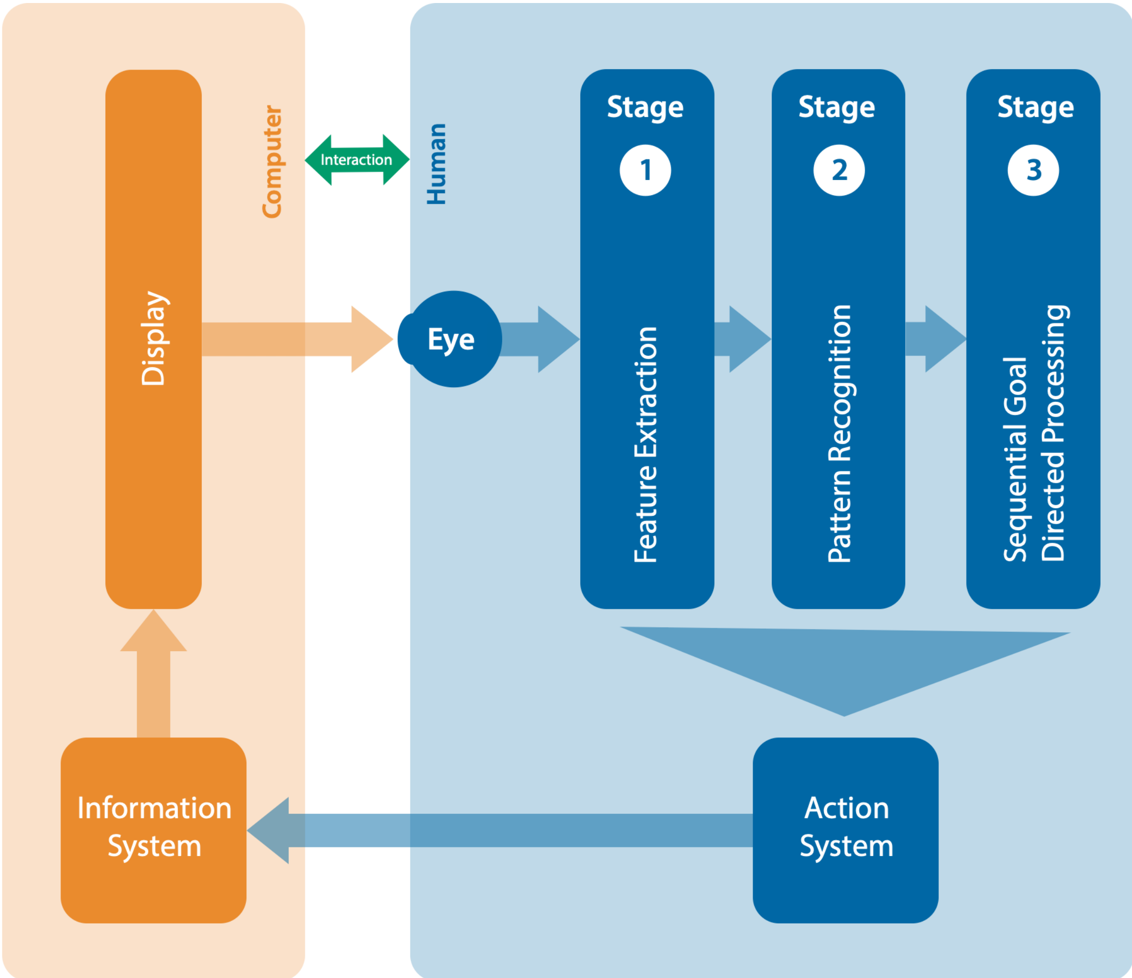


Figure 3-23: A simplified version of Ware's (2004, p.21) three-stage model of vision laid out to map to the basic model of human-computer interaction shown in Figure 3-21

Ware's model provides a useful structure for the description of vision given earlier in section 3.5 and importantly does so in the context of human-computer interaction. This staged model mirrors Marr's (1978; 1982) seminal work on vision and the earlier description of visual search. The stages shown in Figure 3-23 are as follows:

In **Stage One**, billions of neurons work rapidly in a parallel process to extract features from the signal generated by the retina. This information is held briefly and does not require the attention of the user. Ware points out (Ware, 2004, p.21) that if you want the user to *"understand information quickly"* then it is important that things are presented in a way that can be detected effectively at this stage. This equates to the bottom-up guidance described earlier in section 3.5 concerning visual search.

In **Stage Two**, the features extracted in Stage One are used to deduce simple patterns such as contours and regions of the same colour or texture. In addition, it also extracts patterns of motion. This process is not only influenced by the vast amount of bottom-up data from the feature extraction but also top-down by the goals associated with the attention of the user. The bottom-up and top-down nature of the processing is something that is also highlighted in Wicken's (2013) model.

In **Stage Three**, objects are derived from the previous stages and are held in visual working memory by the attention of the user. Ware (Ware, 2004, p.4) also draws out the social context of the goals of the user which is consistent with Bevan's (1995) context of use model which will be discussed later. Stage Three is not the end of the process but leads to other sub-systems such as object recognition, linking to verbal linguistics to allow objects to be named and sub-systems related to motor systems for directing physical movement.

These models show that vision can be represented as a modular and networked system that can be described in stages, with top-down and bottom-up aspects to visual processing. Critically for this research, these models can be directly related to the process of human-computer interaction. What is also apparent is that vast amounts of data are processed and abstracted into objects relevant to the user's goals. This is consistent with the spotlight model described earlier and helps explain phenomenon such as inattentional blindness (the Gorilla Experiment) and how things can be 'missed' or 'misunderstood'.

3.6.3 What about affordances?

Having established that a psychophysical model of vision can be integrated into a view of human-computer interaction it is important to address the topic of 'affordances' which is very closely related to the discussion around visibility so far. This has been deliberately separated from the discussion so far, as it is not only important but also confusion has arisen about the use of the term. The briefest of historical views of the debate helps explain how it has been used and misused, and importantly how it can best be used now. The original term was coined by Gibson (1966), which he then developed (1979) to form the 'direct perception' model of vision. Again, this has been deliberately avoided and only the indirect model discussed so far with regard to vision, this is based on the judgement that the indirect model is the prevailing view that underpins most of what has been discussed so far. It is a point of note that the direct model has merit but is not something that will be elaborated here. However, future work could consider it. The term 'affordance' was then appropriated by Norman (1988) for use in the context of design. Norman (1999) later clarified that it had been misunderstood and that the term *"perceived affordances"* would be better. As the confusion grew, he went on to replace it with a new term, *"signifier"* (Norman, 2011) and then summarised the whole journey (Norman, 2013). This is of note as most of these references are popular books and therefore this changing terminology has gone out to a wide audience and importantly this audience is likely to be one relevant to this research. It is therefore likely that different practitioners will use different words and anecdotally evidence suggest that the word 'affordance' is the dominant one, despite being superseded.

Norman's (2013) summary of the journey and the underlying concepts are helpful. He points out that 'affordances' as defined by Gibson (1966; 1979) are concerned with the 'relationship' between a physical object and person, or more broadly speaking any living thing. As it is relational it is dependent on both the properties of the object and the abilities of the person interacting with it. Designers have their focus on the properties of the 'thing' they are designing. These can get confused as affordances in the

Gibsonian sense of the term. What Norman is trying to encourage is that a design conveys or 'signals' clearly what an object can do and hence 'signifier' is a better term to distinguish it from a pure Gibsonian view of affordance. The concept of signifiers relates very closely to the view of visibility discussed so far. At a practical level 'signifier' is the preferable term but it may be confused with 'affordance' by practitioners. However, what ultimately matters is the underpinning conceptual understanding that helps inform the design of the product in question, in particular relating to the relational and cognitive dimensions of visibility.

3.7 What is Usability? (SRQ 1&2a)

Having established the fundamental nature of the interaction between humans and computers it logically follows to explore the quality of this interaction. In other words what defines how easy or difficult the interaction is, its usability. However, defining how easy something is to use is not easy and proves to be elusive in practice (Hornbæk, 2006). This is summed up well by Gray & Salzman (1998) who liken defining usability to nailing jelly to a wall. Carroll (2010) provides a detailed critique of defining HCI as a whole and he argues that narrow positivist approaches will fail to account for the inherent complexity and multifaceted nature of humans interacting with computers. This is aligned with Gray and Salzman (1998) who see it as a multidimensional problem.

Bevan (1995) provides a useful model that shows a user interacting with a product in the context of the technical, physical and social/organisational environments. The model relates this to measures of effectiveness, efficiency and satisfaction. These three attributes are at the core of the ISO 9241-100 (2010) standard on usability. Although this is a good starting point it does not fully capture all the dimensions of usability. Several efforts have been made to extend these dimensions and key attempts at doing this are shown in Table 3-7 split across both of the following pages. The attributes that are the same or similar are grouped by column. With the final column showing ones that are unique to the particular paper cited.

Table 3-7: A comparison of different sets of usability attributes

Author(s)	Effectiveness	Efficiency	Satisfaction	Learnability
Shneiderman (1983b)		Speed of task performance	Subjective user satisfaction	Time for users to learn specific functions
Shackel (1991)	Effectiveness		Attitude	Learnability
Nielsen (1993b, p.26)		Efficiency	Satisfaction	Learnability
ISO 9241-100 (2010)	Effectiveness	Efficiency	Satisfaction	
Constantine & Lockwood (1999)		Efficiency in use	User Satisfaction	Learnability
Abran et al. (2003)	Effectiveness	Efficiency	Satisfaction	Learnability
Seffah (2006)	Effectiveness	Efficiency	Satisfaction	Learnability
Rubin et al. (2008, p.16)	Effectiveness	Efficiency	Satisfaction	
Harrison et al. (2013)	Effectiveness	Efficiency	Satisfaction	Learnability

Table 3-7 continued

Memorability	Errors	Safety	Unique to author(s)
User retention of commands over time	Rate of errors by users		
			Flexibility
Memorability	Errors		
Rememberability		Reliability in use	
		Security	
		Safety	Productivity, Trustfulness, Accessibility, Universality, Usefulness
			Usefulness, Accessibility
Memorability	Errors		Cognitive Load

Grudin (1992) draws the important distinction between 'utility' and 'usability'. He points out that something can be potentially useful (utility) but unusable (usability) or conversely something can be usable but offer little or no value (utility). Seffah (2006) and Rubin and Chisnell (2008, p.16) include 'usefulness' in their respective sets of attributes which equates to 'utility'.

A view of these different usability properties in Table 3-7 and Bevan's (1995) model show that usability is an outcome of a user interacting with a product to achieve their goals within a certain context. The attributes of the product and indeed the nature of the context contribute towards it being a successful one. This perspective is unpacked in detail by Bevan, Kirakowski, and Maissel (1991). However, the 'cause' (product attributes and context) can be conflated with the 'effect' (usability). Examples of this conflation exist in the literature such as Ham (Ham, 2013) who describes usability as a quality attribute of the system.

So, usability is:

1. Not easy to define
2. Multidimensional
3. Context-dependent
4. An outcome (cause [product attributes] and effect [outcome of use] can be confused)
5. Distinct from utility

In terms of a pragmatic definition, the ISO 9241-100 (2010) is a good baseline.

> *"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use."*

A substantially extended view of this can be provided by developing Bevan's (1995) basic model to include an expanded set of attributes and also includes utility (described as task match) by comparing the user's task goals to the actual output from the use of the product or system. This is shown in Figure 3-24.

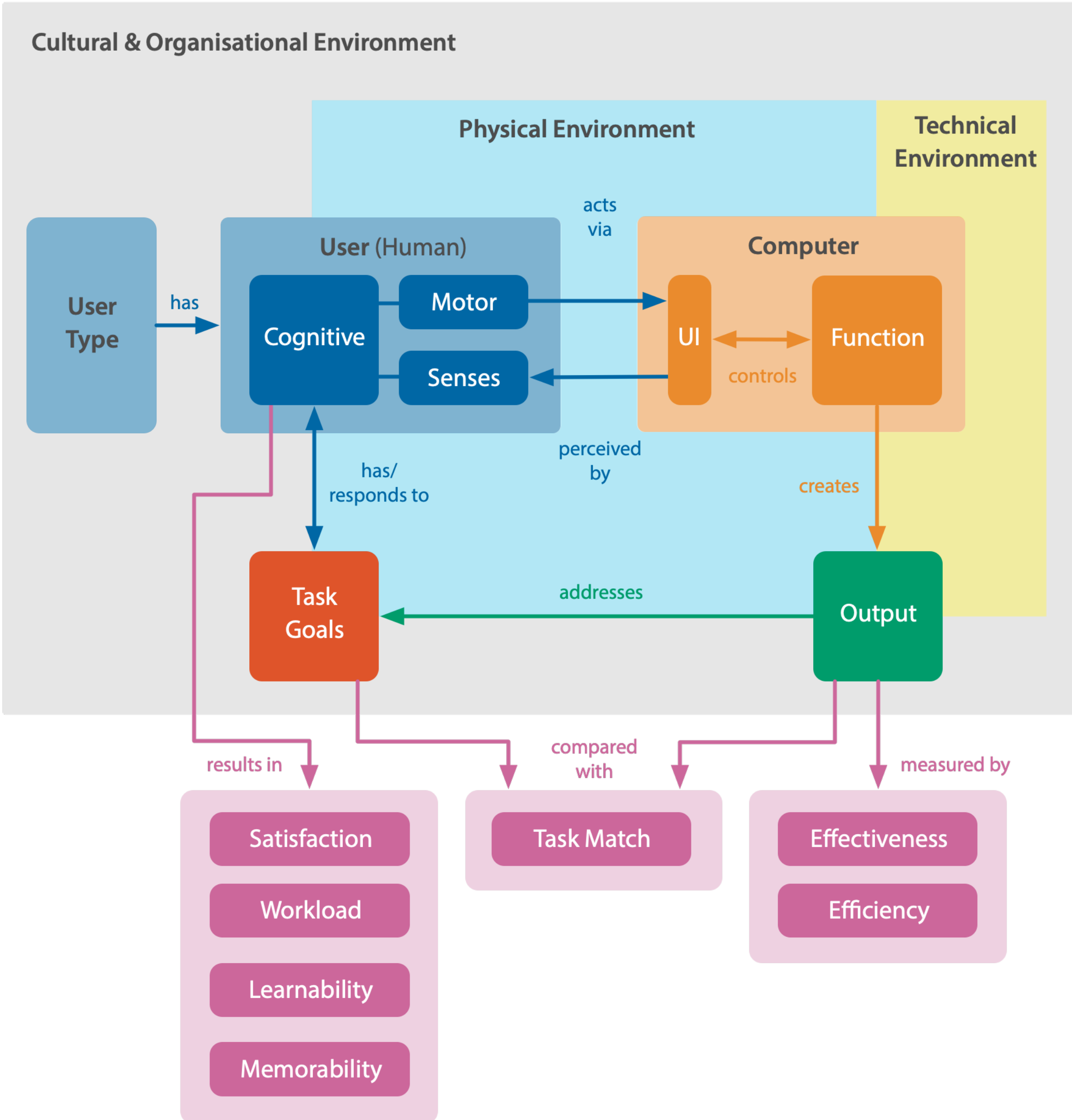


Figure 3-24: An extended model of usability based on Bevan (1995)

Having established what a human-computer interaction is and a definition of the quality of the interaction in terms of usability, the next question is what types of 'usability problems' are there that lead to poor outcomes.

3.8 What is a Usability Problem? (SRQ 1&2b)

Not unsurprisingly definitions of what a 'usability problem' is, like 'usability', prove difficult. Lavery, Cockton, and Atkinson (1997), as well as Manakhov and Ivanov (2016), provide reviews of different definitions of what a 'usability problem' is. The former's definition provides a good match with the previous section's definition of usability i.e. one that is outcome-based. The definition is:

> *"A usability problem is an aspect of the system and/or a demand on the user which makes it unpleasant, inefficient, onerous or impossible for the user to achieve their goals in typical usage situations."*

They also provide a model that breaks down a usability problem into different elements. These are shown in Figure 3-25 below.

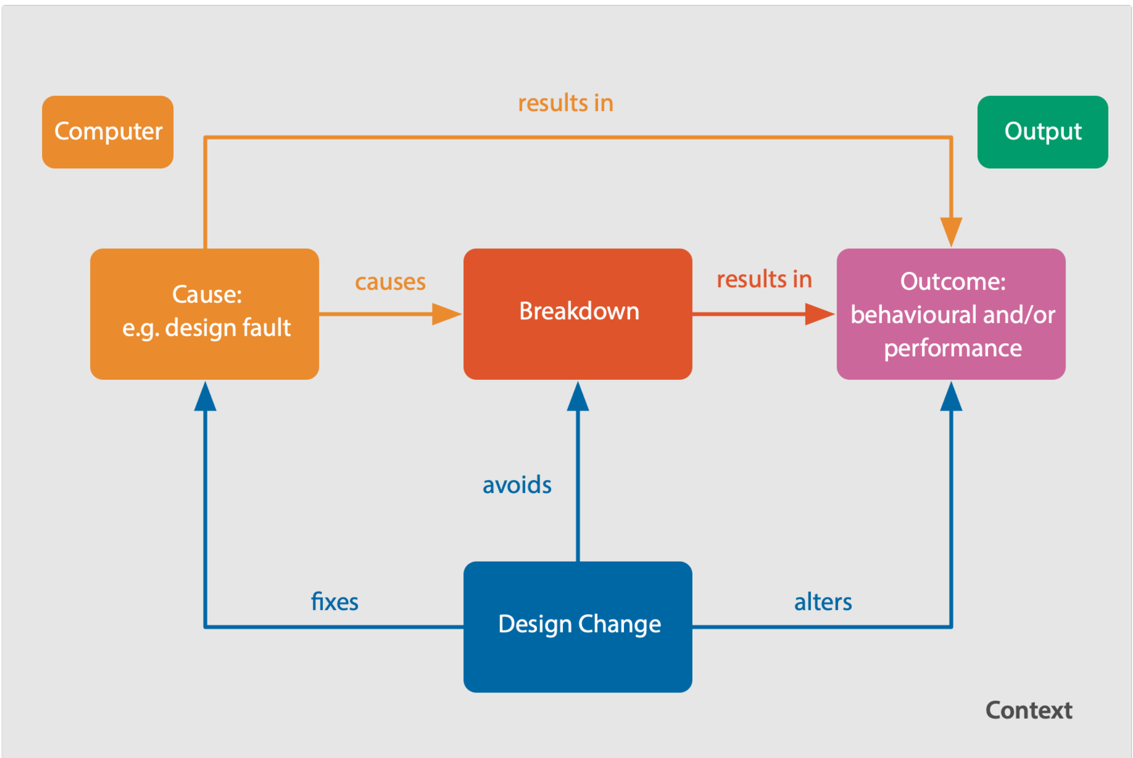


Figure 3-25: Generic model of a usability problem based on Lavery, Cockton, and Atkinson (1997) with the addition of 'Computer' and 'Output' labels to show mapping to the usability model presented in figure 3-21

This model is helpful as it not only articulates the fundamental cause and effect nature of a usability problem but by breaking it down into a series of components it highlights how descriptions can vary depending on the focus of that description e.g. it could be 'cause' centric or 'output' centric. This complexity is consistent with the problem of poor inter-rater reliability in usability studies that has been called the 'evaluator effect' by Jacobsen, Hertzum, and John (1998) and is well attested to (Vermeeren, van Kesteren and Bekker, 2003; Hertzum and Jacobsen, 2003a; Hornbæk and Frøkjær, 2008) and persists as a problem (Hertzum, Molich and Jacobsen, 2014; Sauro, 2014). The studies show that the problem exists for both novice and expert evaluators and pervades identification, severity rating and description. Attempts have been made to make the description process more rigorous and consistent, such as Capra's (Capra and Smith-Jackson, 2005; Capra, 2007) ten guidelines on usability problem reporting.

The process of problem identification and description via usability evaluation methods (UEM's) is addressed in the next section and the issues raised above became a reality in this research as part of the empirical work, which is discussed later (see Section 6.7.1).

In addition to a conceptual understanding and the associated empirical issues of the evaluator effect, it is also useful to consider the role of visibility in usability problems. Keenan, Hartson and Kafura (1999) provide a categorisation of usability problems types known as the Usability Problem Taxonomy (UPT). This is shown in Figure 3-26. This was derived from the iterative analysis of 406 real-world usability problems which were based on 645 usability problem descriptions from 5 different projects. A problem can be described in terms of its artefact and task component although partial classification is also allowed.

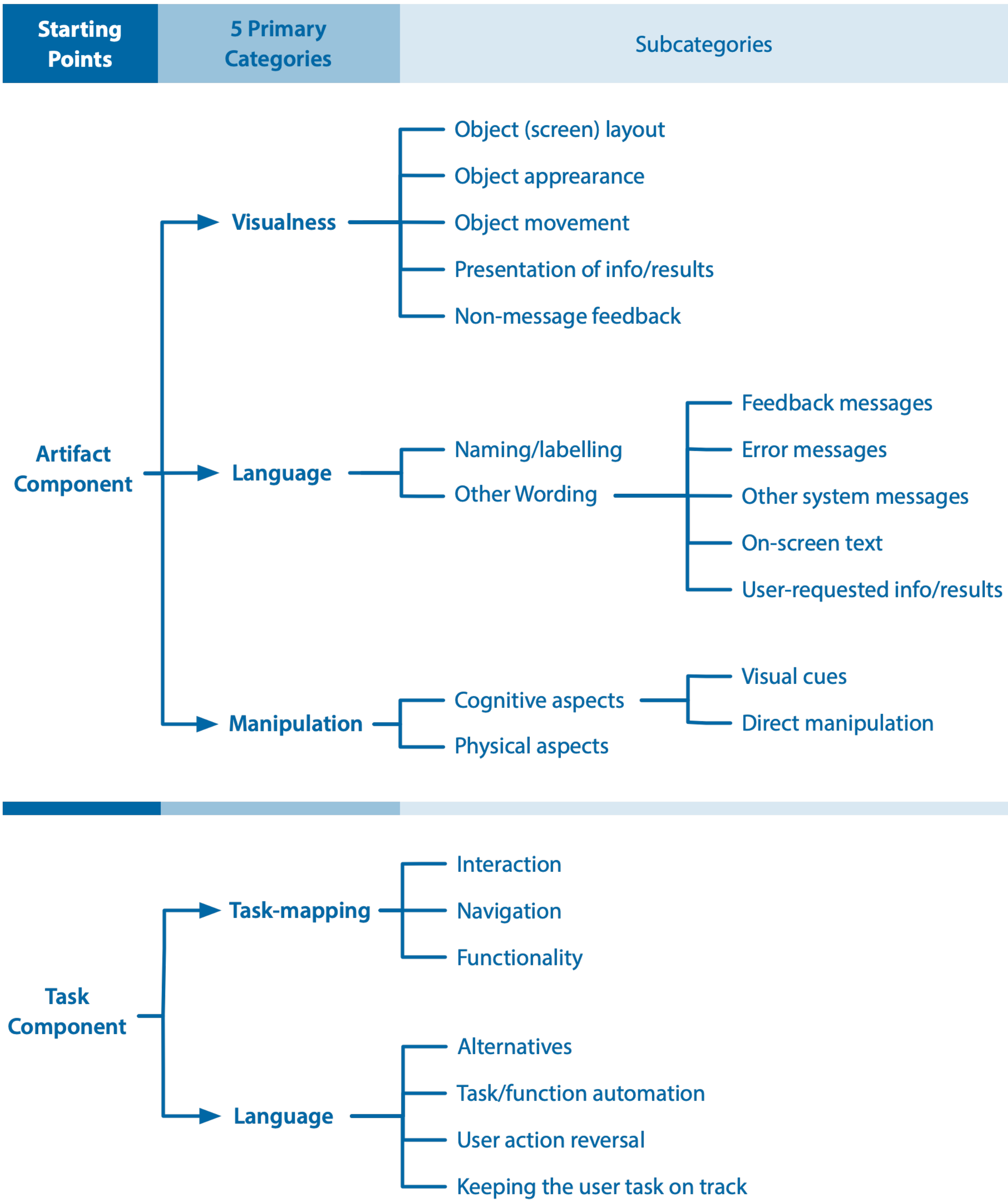


Figure 3-26: Usability Problem Taxonomy categories redrawn from Keenan, Hartson and Kafura (1999)

The subcategories are stated as being mutually exclusive, although the inclusion of 'visual cues' under 'cognitive aspects' is confusing and not explained. However, the primary category of 'visualness' is consistent with this research. Also, a study of 20 randomly selected problems that were analysed by 7 participants led to 'visualness', at 30%, being the highest incidence of classified artefact problems. This compared to 23% for 'language' and 11% for 'manipulation'. These findings address SRQ 2b, concerning whether the problems caused by visibility are significant, as visibility is a key part of the UPT model and the lead cause of problems in the sample analysed.

Null classification counted for the remainder of the problems analysed, which raises concerns about its effectiveness in categorising problems. However, the paper does address this through discussion of a range of potential underlying issues, such as poor problem descriptions and the lack of contextual information. So, caution is required in using it as evidence, but none the less it is a substantive piece of work and consistent with the direction and findings of this research. It is also worth noting that the UPT was built on the User Action Framework (Andre et al., 2001). This draws on Norman's (1986) Stages of Action and is consistent with Neisser's (1976) Perceptual Cycle.

Having considered the nature of usability problems this leads to considering evaluation methods that can help identify them.

3.9 How is usability evaluated? (SRQ 1&2c) and how well do these approaches address visibility (SRQ 1&2d)

Usability evaluation can be seen as having two key aims concerning the 'identification' and 'quantification' of usability problems (Hornbæk, 2006; Clarkson, Waller and Cardoso, 2015). Identification is key to being able to address problems to improve performance. Quantification is key in qualifying associated risks and prioritising what issues need fixing the most (Nielsen, 1994, p.103; Karat, 1997).

Quantification comes in a variety of different forms from the assessment of potential impact, to the likelihood of occurrence (Nielsen, 1994, p.104). A variety of severity scales have been devised (Nielsen, 1994, p.103; Dumas and Redish, 1999, pp.323-324; Rubin and Chisnell, 2008, p.262). These typically have an endpoint of unusable or similar. The notion of usable, or more specifically unusable, has been extended to the concept of exclusion within inclusive design. This resulted in an evaluation approach that can estimate the proportion of the population that would be unable to use a product or an aspect of its functionality (Clarkson, Waller and Cardoso, 2015). This quantification is valuable in raising the profile of problems within an organization and helping prioritise what problems should be addressed.

Such rating mechanisms and estimates of exclusion are complemented by more objective measures based on time and task completion (Shneiderman, 1983b; Shackel, 1991; Wixon and Wilson, 1997, p.666). Also, the number of errors and the error rate can be considered. These are shown in Table 3-8 below concerning the usability attributes described in the previous section.

Table 3-8: Specific measures of usability attributes

Usability Attribute	Measures
Effectiveness	Percentage of a defined task completed at or above the required level of performance (where performance can be considered in part by errors) by specified users in a specified context Task completion compared to an expert
Efficiency	Task completion time (or more broadly resource expenditure) Task completion time compared to an expert
Satisfaction	Satisfaction questionnaires
Learnability	Time to reach a specified level of effectiveness and efficiency
Errors	Number of errors Error rate Time to recover
Memorability	Effectiveness and or efficiency after a specified delay in use

The National Physical Laboratory in the UK (Bevan, 1995) took these measures further through the development of a set of usability metrics. With the following measures:

$$Task\ effectiveness = \frac{Quantity \times\ Quality}{100}\%$$

In task effectiveness 'quantity' refers to the number of sub-tasks completed, and 'quality' refers to the quality of the output. For example, if the user's task is 'entering text' and there are minor errors the task may be deemed to be completed (quantity) and someone can correctly understand the text, but the quality is not as high as it could be i.e. there are typos. The quality measure can become quite subjective but allows for the variability of user outcomes rather than a simple pass or fail.

From 'task effectiveness' it is possible to calculate 'temporal efficiency' as follows:

$$Temporal\ Efficiency = \frac{Effectiveness}{Task\ Time}$$

If the 'temporal efficiency' has also been measured for an expert user, it is possible to calculate the 'relative user efficiency' compared to this expert as follows:

$$Relative\ User\ Efficiency = \frac{User\ Efficiency}{Expert\ Efficiency} \times\ 100\%$$

This measurement acts as a measurement of learning if repeated over time. Finally, it is possible to calculate the productive period as follows:

$$Productive\ period = \frac{Task\ Time - Unproductive\ Time}{Task\ Time} \times 100\%$$

Bevan (1995) describes how such effectiveness and efficiency measures can be complemented by subjective measures of satisfaction. Table 3-9 shows a range of such measures that cover perceived usability. They vary in terms of the number of items and scope. The NASA TLX, for example, has a narrower focus on subjective workload. The QUIS is broader in scope and explicitly addresses 'screen' as a factor covering a number of visibility issues, see Figure 3-27 for a screenshot of the electronic data entry form.

Table 3-9: A summary of different user assessment instruments

Scale		No items	Reference
QUIS	Questionnaire for User Interaction Satisfaction	34 (v 7.0)	(Chin, Diehl and Norman, 1988; Norman, Shneiderman, Harper and Slaughter, 2007)
EUCS	End-User Computing Satisfaction	18	(Doll and Torkzadeh, 1988)
ASQ	After Scenario Questionnaire	3	(Lewis, 1991)
SUMI	Software Usability Measurement Inventory	50	(Kirakowski and Corbett, 1993; van Veenendaal, 1998)
PSSUQ/CSUQ	Post-Study System Usability Questionnaire	19	(Lewis, 1995)
SUS	System Usability Scale	Comprehensive (10 questions)	(Brooke, 1986)
USE	Usefulness, Satisfaction, and Ease of use	30 items	(Lund, 2001; Faria, Pavanellie and Bernardes, 2016)
NASA TLX	Task Load Index	6 items	(Hart and Staveland, 1988)

PART 2: Screen

2.1 Characters on the computer screen	hard to read	easy to read	1 2 3 4 5 6 7 8 9	NA
2.1.1 Image of characters	fuzzy	sharp	1 2 3 4 5 6 7 8 9	NA
2.1.2 Character shapes (fonts)	barely legible	very legible	1 2 3 4 5 6 7 8 9	NA
2.2 Highlighting on the screen	unhelpful	helpful	1 2 3 4 5 6 7 8 9	NA
2.2.1 Screen layouts were helpful	never	always	1 2 3 4 5 6 7 8 9	NA
2.2.1.1 Amount of information that can be displayed on screen	inadequate	adequate	1 2 3 4 5 6 7 8 9	NA
2.2.1.2 Arrangement of information on screen	illogical	logical	1 2 3 4 5 6 7 8 9	NA
2.2.2 Sequence of screens	confusing	clear	1 2 3 4 5 6 7 8 9	NA
2.2.2.1 Next screen in a sequence	unpredictable	predictable	1 2 3 4 5 6 7 8 9	NA
2.2.2.2 Going back to the previous screen	impossible	easy	1 2 3 4 5 6 7 8 9	NA
2.2.2.3 Progression of work related tasks	confusing	clearly marked	1 2 3 4 5 6 7 8 9	NA

Please write your comments about the screens here:

____

____

____

____

Figure 3-27: Image capture of QUIS 7.0 questions regarding screen issues (Norman et al., 2007)

QUIS is a mature instrument, it is on its seventh version, and highlights the importance placed on the visual, but does not directly address the three M's (missing, missed and misunderstood).

User questionnaires are only one form of usability evaluation method (UEM). Ivory & Hearst (2001) provide a useful categorization of different UEM types. Table 3-10 has a selection of methods based on this categorization. The selection is based not only on the literature already discussed but also on two comprehensive reviews of UEM's by Fernandez, Insfran and Abrahão (2011) and Rajeshkumar, Omar and Mahmud (2013), as well as the detailed work on UEM's by Gray and Salzman (1998). In addition, the perspective of what is taught from an undergraduate HCI book (Rogers, Sharp and Preece, 2011) was considered to try and represent what a typical usability practitioner might have been taught to use. Finally, a regulatory viewpoint of UEM's for medical device usability evaluation was considered as this is key in medical device development and therefore representative of what practitioners working in this domain will do in practice (CENELEC, 2015).

Table 3-10: Summary of key usability evaluation methods

Category	Example	Visibility Coverage
Testing		
	Lab-based performance measurement	The measures outlined in Table 3-8 do not address visibility directly
	Logging based analysis	Typically, user events such as mouse clicks or data entry i.e. not visibility related. Can include eye-tracking but this only addresses what a user looks at and not why or what they have perceived (Hilbert and Redmiles, 2000).
Inspection		
	Cognitive walkthrough	Visibility is not an explicit part of the evaluation and therefore assessor dependent
	Heuristic evaluation	Includes holistic visibility heuristics
	Expert review	Visibility is not an explicit part of the evaluation and therefore assessor dependent
	Standards or guidelines-based inspection	Specific and typically narrow threshold-based visibility measures

Table 3-10: continued

Category	Example	Visibility Coverage
Inquiry		
	Questionnaire	Specific vision questions in QUIS but not in many of the other similar questionnaires
	Interviews	Highly assessor dependent
	Focus groups	Highly assessor dependent
	Diaries	User dependent and likely to have limited expertise in user interface issues
Analytical modelling		
	GOMS 'Family'	The different variants of GOMS do include a perceptual element and recognize that the processing times increase with the *"complexity of the signal being perceived"*, however such modelling does not directly address the broader visibility issues discussed (John and Kieras, 1996).
	Guided Search	Limited to search and laboratory tests (Wolfe, 2007)
Simulation		
	ACT-R including PAAV and SNIF-ACT vision modules	ACT-R includes several vision modules from the specific (SNIF-ACT) to the more general PAAV but are very complex and not currently suitable for typical user interface practitioners (Emond and West, 2004; Katsanos, Tselios and Avouris, 2010; Nyamsuren and Taatgen, 2013).
	ISAR (Information scent absorption rate)	ISAR and a number of similar measures make use of the ACT-R information scent prediction model (SNIF-ACT) and are focused primarily on web browsing i.e. information search and the links that have to be followed (Katsanos, Tselios and Avouris, 2010)

Table 3-10 includes an overview of how different UEM's cover visibility. In the case of standards and guidelines, this is broken down further as many contain explicit guidance on visibility. NISTIR 7889 (2014) provides a useful comparison of different standards and their content. A selection of these is shown in Table 3-11 that were reviewed for their approach to visibility. These range from general guidance to the detailed specification of contrast ratios, use of colour and the size of text and objects. Size is normally specified in terms of the angle subtended to the eye to allow for different viewing distances. Such an approach maps to the 'window of visibility' discussed earlier. Indeed, the NASA SP-2010-3407 Human Integration Design Handbook (Runco and Scott, 2014) has a whole section explaining how the visual system works and the expression 'visibility window' came from researchers in the vision group at NASA. The standards listed in Table 3-11, particularly the MIL-1472G (US Department of Defense, 2012) and NASA/SP-2010-3407 are based on decades of expertise, tie into the research literature and in the case of NASA have contributed to it. Therefore, these represent a key resource albeit optimized to specific application domains, for example, NASA has to deal with the extremes of gravitational pull and vibration associated with space flight.

Table 3-11: A selection of human factors standards across different sectors that address visibility

Standard	Organisation/Ref	Name	Sector
MIL-1472G	U.S. Department of Defense (2012)	Design Criteria Standard: Human Engineering	Defence
NASA/SP-2010-3407/REV1	National Aeronautics and Space Administration (2014)	Human Integration Design Handbook (HIDH)	Space
DOT/FAA/AM-01/17	U.S. Department of Transportation Federal Aviation Administration (2001)	Human Factors Design Guidelines for Multifunction Displays	Aviation
No reference number	Food and Drug Administration (2016)	Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff	Medical
IEC 62366-1:2015	European Committee for Standardization (2015)	Part 1: Application of usability engineering to medical devices	Medical
IEC TR 62366-2:2016.	European Committee for Standardization (2016)	Medical devices, Part 2: Guidance on the application of usability engineering to medical devices	Medical
NISTIR 7889	National Institute of Standards and Technology (2014)	Human Engineering Design Criteria Standards Part 1: Project Introduction and Existing Standards	National Security

Standard	Organisation/Ref	Name	Sector
EN ISO 9241-303:2011	International Organization for Standardization (2011)	Ergonomics of human-system interaction - Part 303: Requirements for electronic visual displays.	General VDU use
EN ISO 9241-161:2016	International Organization for Standardization (2016)	Ergonomics of human-system interaction - Part 161: Guidance on visual user-interface elements.	General VDU use
EN ISO 9241-112:2017	International Organization for Standardization (2017a)	Ergonomics of human-system interaction - Part 112: Principles for the presentation of information.	General VDU use
EN ISO 9241-125:2017	International Organization for Standardization (2017b)	Ergonomics of human-system interaction Part 125: Guidance on the visual presentation of information.	General VDU use
EN ISO 9241-306:2018	International Organization for Standardization (2018)	Ergonomics of human-system interaction - Part 306: Field assessment methods for electronic visual displays.	General VDU use

Overall standards that do address visibility tend to take a narrow, prescriptive approach, which understandably helps with testing for compliance but does not address the wider issues associated with visibility that have been described. The QUIS Questionnaire (Norman et al., 2007) discussed earlier takes a similarly narrow approach albeit on a self-report basis.

Other inspection based UEM's such as Nielsen and Mack's (1994) heuristic evaluation take a broader more open approach, consistent with the notion of heuristics. A selection of three of the ten heuristics that emphasize the importance of visibility is shown below.

> *"1.* **Visibility** *of system status:*
>
> *The system should always keep users informed about what is going on, through appropriate feedback within reasonable time."*
>
> *"6. Recognition rather than recall:*
>
> *Minimize the user's memory load by making objects, actions, and options* **visible***. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be* **visible** *or easily retrievable whenever appropriate."*
>
> *"8. Aesthetic and minimalist design:*
>
> *Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative* **visibility***."*

These heuristics, unlike the narrower view of the standards, do represent a broader view of visibility. However, they do not address the 3 M's explicitly and are heavily reliant on the skill and knowledge of the assessor.

Cognitive modelling tools such as ACT-R (Nyamsuren and Taatgen, 2013) are developing increasingly more sophisticated and general-purpose vision elements to their models. However, they are complex, and it can be argued not currently suited to your typical usability practitioner.

A basic characterization of the situation indicates that some UEM's do address visibility directly but tend to be either too narrow and prescriptive or open and heavily reliant on the knowledge of the assessor. None of them explicitly address the visibility issues covered by the 3 M's Model. This leaves a significant gap in terms of evaluation which could be addressed either by a new standalone approach or augmenting existing methods to take into account the 3 M's visibility issues.

This leads to unpacking the third Initial Research Question which is:

> **IRQ 3:** *What can be done to improve it?*

This can be broken down as follows, firstly, can a broader form of visibility be framed from a cognitive perspective. The word 'cognitive' is used here to distinguish it from a narrower psychophysical perspective that looks at variability around a threshold of visibility. The question is as follows:

> **SRQ 3a:** *Is a cognitive-based framework for understanding and representing the visibility of user interfaces more effective than one based on a simple visibility threshold?*

For such a framework to be useful to practitioners leads to the question of whether it can be embodied in an evaluation tool, which was highlighted as a gap in current evaluation methods. The question is worded as follows:

> **SRQ 3b:** *Can such a framework be embodied in an evaluation tool that predicts more usability problems than current approaches?*

Finally, if such a tool is possible then will it be used by usability practitioners to improve the visibility of user interfaces. This question is worded as follows:

> **SRQ 3c:** Can and will practitioners use such a tool to reliably improve the visibility of user interfaces?

This leads to the complete set of research questions shown in figure 3-28 below.

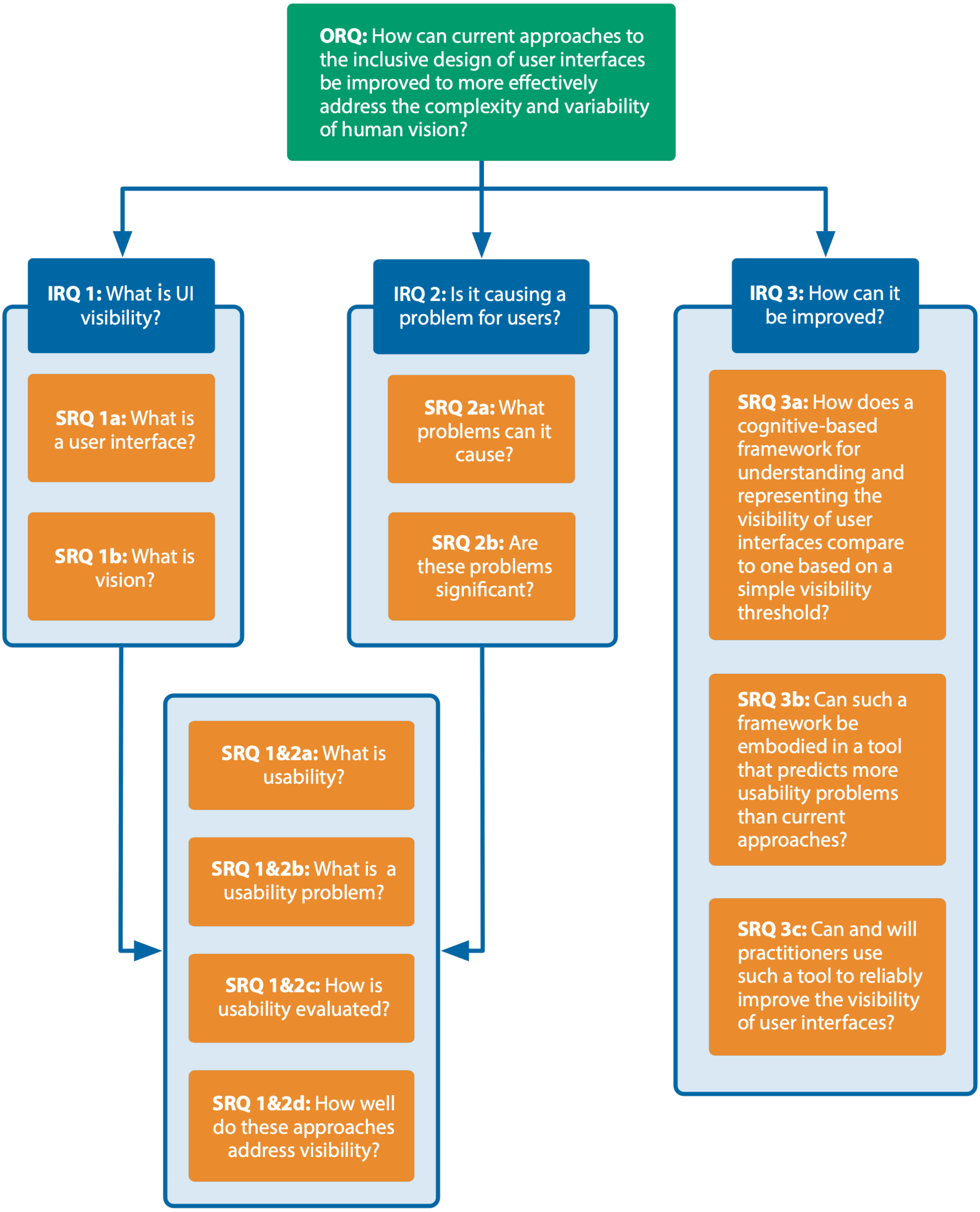


Figure 3-28: The complete set of research questions

3.10 DRM Reference and Impact Models

The DRM makes use of 'reference' and 'impact' models to summarise the existing and preferred situations (Blessing and Chakrabarti, 2009, p.20). The models consist of the 'influencing factors' and the relationships between them. The reference model represents the current situation against which any intended improvements are 'referenced' against. The intended improvements are embodied in the impact model. The reference model for this research is shown in Figure 3-29 and draws together the key elements from this research clarification stage. Filling the gap in usability evaluation methods regarding visibility is shown in the impact model in Figure 3-30 represented by the 'UI Visibility Analysis Tool'. Both models omit the annotation of the links to aid readability.

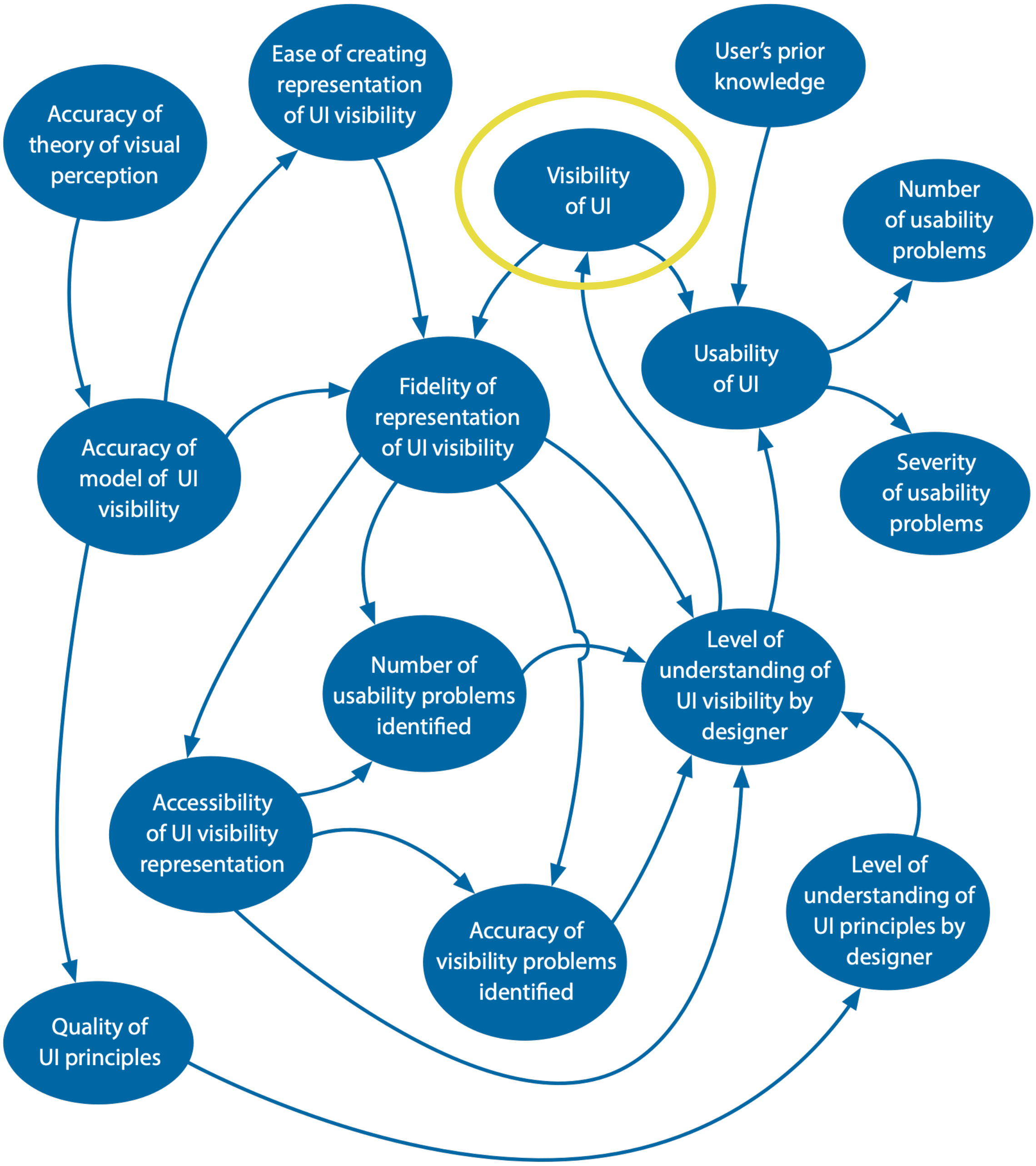


Figure 3-29: DRM reference model with focus areas (Blessing and Chakrabarti, 2009, pp.25-29)

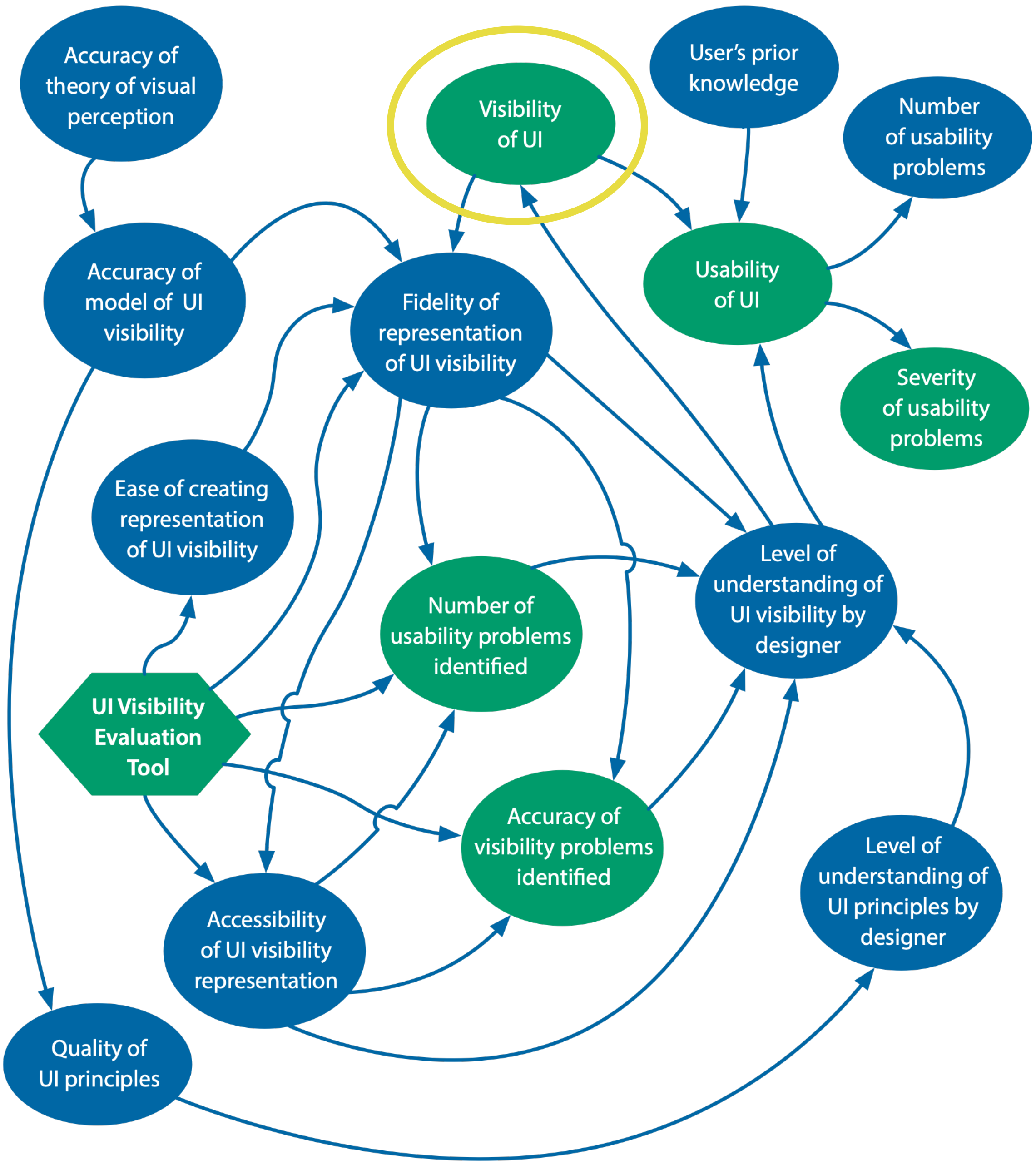


Figure 3-30: DRM impact model with focus areas in green (Blessing and Chakrabarti, 2009, pp.25-29)

3.11 Summary

The literature review, framed by initial empirical work, has produced an integrated conceptual underpinning across a wide range of contributory areas. This has led to the following key outputs:

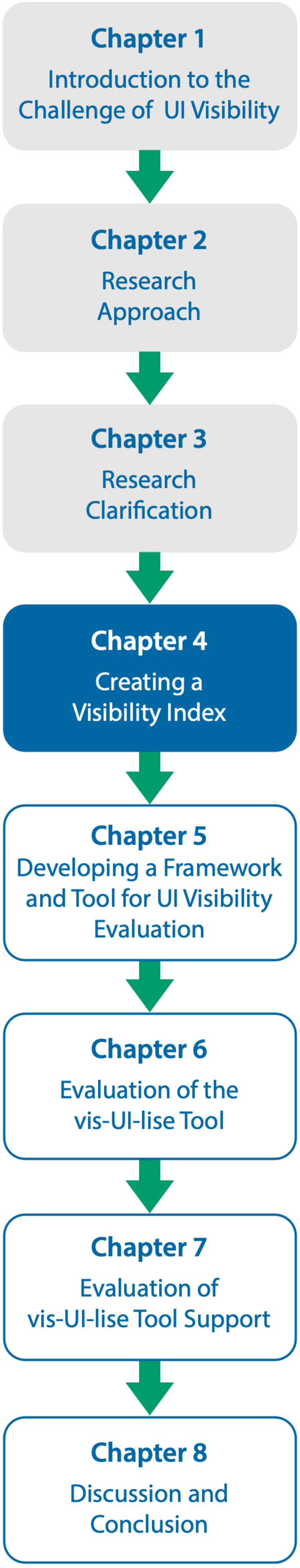
- A model of visibility problems covering visual elements that are missing, missed or misunderstood in a model known as the 3 M's. This represents the key underlying problem that the research aims to address.
- An understanding of vision as an attentional, zooming spotlight
- Both the notion of the spotlight and the criticality of context in visibility have a large bearing on visual performance
- This perspective of vision is contrasted to a naïve model of vision which considers it as producing a uniform image across the visual field. This potential misunderstanding amongst practitioners might cause problems during the design process.
- A comprehensive model of usability was developed from the literature and when combined with the broader view of visual performance led to the following definition of user interface visibility:

  UI Visibility concerns the demands the user interface's visual properties put on the user's visual ability in the context of use (task) to produce the desired outcome (goals) in a physical, technical & social environment, where visual ability is considered from a broader cognitive perspective including the user's prior knowledge.

- Having established an understanding of UI visibility, current usability evaluation methods were reviewed with regard to their support for assessing visibility. The conclusion from this is that current UEM's do not support this holistic view of UI visibility in a manner suitable for usability practitioners.

- Finally, a fully elaborated set of research questions was devised that underpin the rest of the research with a focus on developing support for usability evaluation, either in the form of a new method or support for existing ones.

Having identified three key types of visibility issues in terms of the 3 M's the next chapter looks at approaches to 'visualising UI visibility'. In particular, how prevalent the invisibility represented by 'missing' is in modern user interfaces.


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 4

Creating a Visibility Index

4.1 Introduction[1]

The previous chapter highlighted that in Microsoft Word™ the number of commands in Word 1.0™ was about 100 but by Word 2003™ it had exceeded 1500. When Microsoft asked users what they wanted in the next version of Office™, 9 out of 10 asked for features they already had in their current version (Caposella, 2005). Not only is not finding the required function a problem, but the opposite problem of accidentally activating an undesired function is an issue too. For example, Apple's iOS™ has a delete mode for apps that is activated by a 'touch and hold' on an app icon (Apple Inc., 2016b). This can lead to accidental deletion of an app and its data or simply confusing the user.

What we see is the combination of a proliferation of features with interfaces that lack visibility as they move to a minimalist design style. This is understandable as many post-WIMP devices (Nielsen, 1993a; Gentner and Nielsen, 1996; van Dam, 1997; 2001) must work within the constraints of smaller form factors such as those of smartphones or watches. Before addressing the deeper issues of visibility highlighted already, it is prudent to quantifying how many features are available and how many are visible directly to the user and how many are 'missing'. Effectively this creates a simple 'visibility index' for a user interface. It forms an extension to the initial product analysis already conducted and is the detailed analysis described in Figure 3-14 & 3-15 in Chapter Three and is updated in Figure 4-1. As such it represents a continuation of Descriptive Study One within the DRM framework (Blessing and Chakrabarti, 2009, pp.15-16).

[1] A significant proportion of this chapter is based on Hosking & Clarkson (2017) and edited and expanded to integrate with the overall flow of this dissertation.

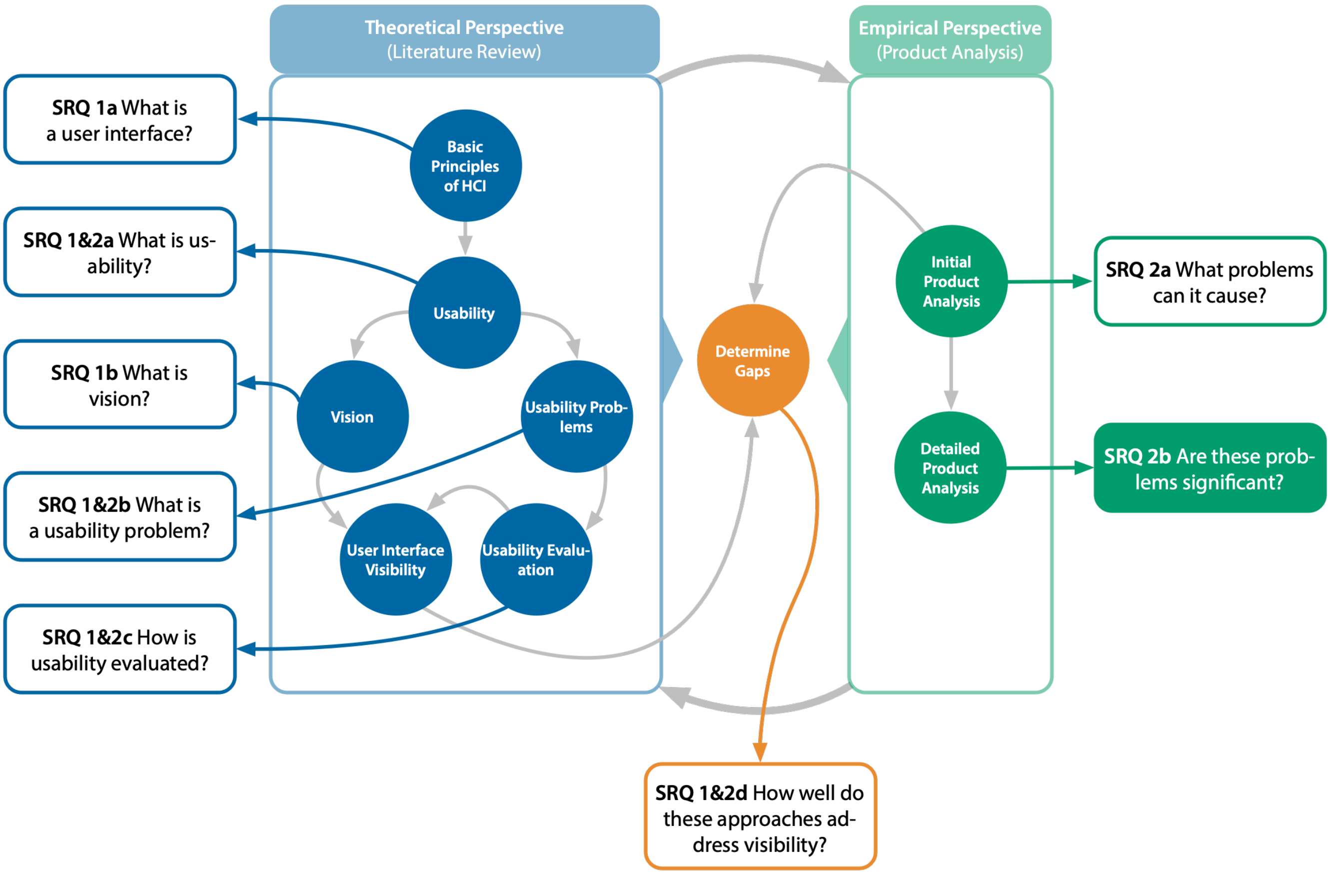


Figure 4-1: Positioning of the detailed product analysis as part of Descriptive Stage One

The first attempt at doing this was for file management on the Apple Macintosh desktop. This was chosen because it represents a widely used graphical user interface and one that has evolved over a long period (English, Engelbart and Berman, 1967; Myers, 1998). This means it is not the only representative of something in everyday use, but it is also an opportunity to explore how things might be changing. Example screenshots taken from two versions are shown in Figures 4-2 and 4-3 where there is a 25-year gap between them [1988 versus 2013].

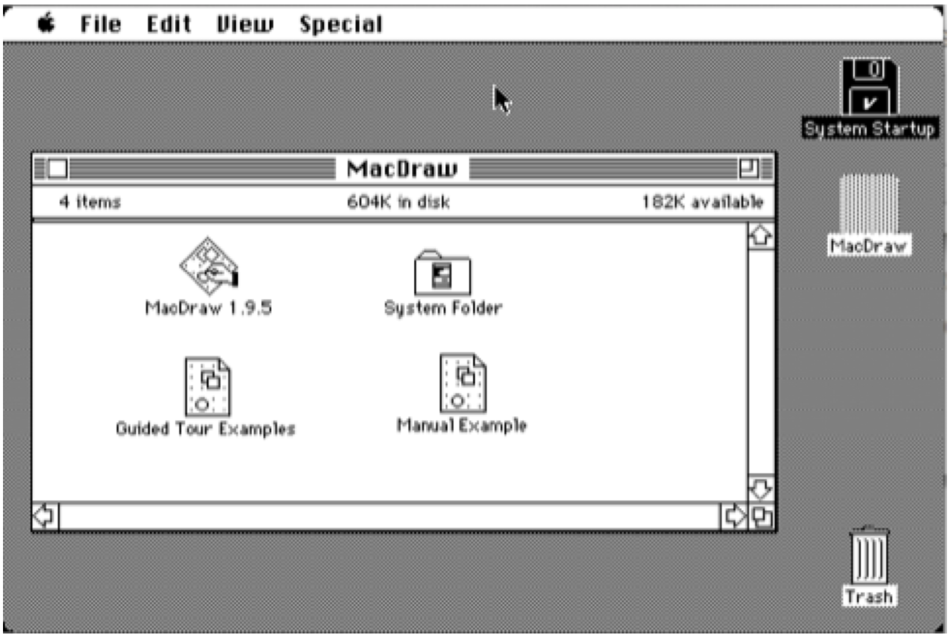


Figure 4-2: Macintosh System Software 6 (Apple Computer Inc., 1988) showing a folder on the desktop (a screenshot produced from a simulator running this version)

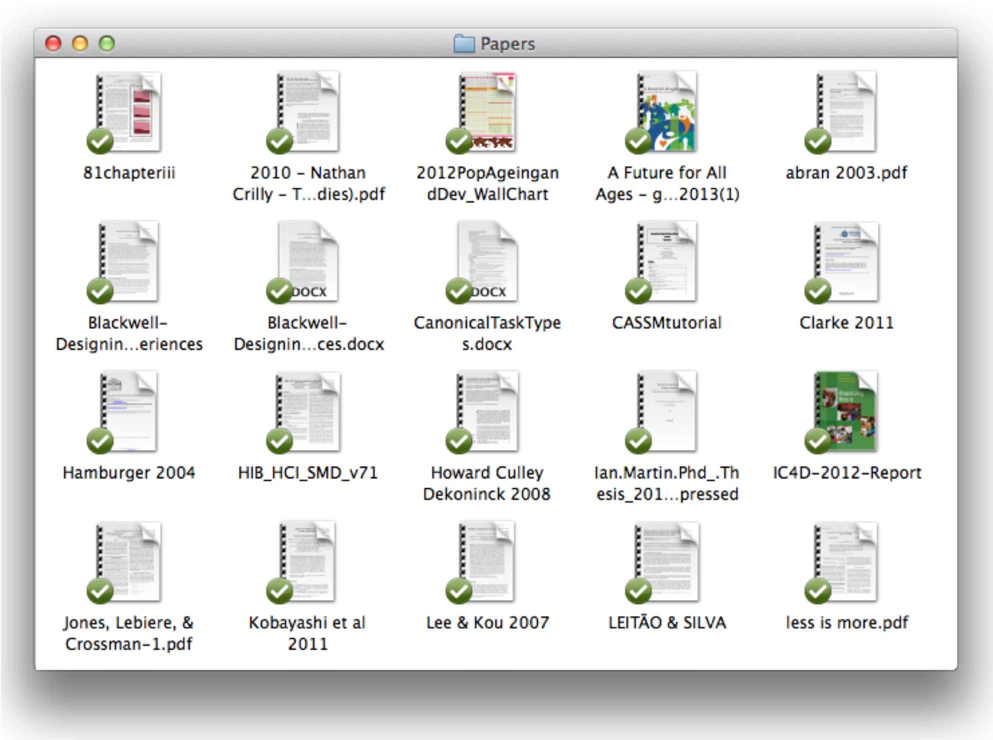


Figure 4-3: Mac OS X 10.9 (Apple Inc., 2013) showing a folder on the desktop (screenshot)

The process consisted of listing all the folder and file management functions. Then each of these functions was assessed in turn for both examples. This is done using a simple visibility rating scale as described in Table 4-1, covering four categories.

Table 4-1: Simple visibility rating scale

Visibility Rating	Description
Yes	The function has a visible control
No	The function has no visible control
Transient	The function becomes visible for example in Mac OS X 10.9 the scroll bar in the window becomes visible then scrolling is initiated or the cursor graphic changes when placed over the window boundary to indicate resizing is possible
Not available	The function is not available in this application

The 'yes' category relates to visible items. It is worth noting that even if something is a 'yes' it could still result in items being 'missed' or 'misunderstood' but this problem is not covered in this basic analysis. The categories of 'no' and 'transient' relate to items that are 'missing' either permanently or on a transient basis. There is a fourth category of 'not available' to cover the comparison of products where a product may not have a particular function that is available in the other product. Using these ratings, it is then possible to calculate the percentage of functions that are always visible (i.e. excluding transient visibility). This is a simple calculation of the functions rated as 'yes' divided by the total number of functions analysed. This analysis is shown in Table 4-2.

Table 4-2: Simple visibility scoring for file management in Macintosh System 6 (Apple Computer Inc., 1988) compared to Mac OS X 10.9 (Apple Inc., 2013)

Function	System 6	OS X 10.9
Close	Yes	Yes
Move	Yes	No
Resize Left	Not available	Transient
Resize Right	Yes	Transient
Resize Top	Not available	Transient
Resize Bottom	Yes	Transient
Maximise	Yes	Yes
Minimise	Not available	Yes
Full Screen	Not available	Yes
Activate item	No	No
Move item	Yes	Yes
Delete item	Yes	Yes
Create Folder	Yes (File menu)	Yes (File menu)
Vertical Scroll	Yes	Transient
Horizontal Scroll	Yes	Transient
Visibility Index	**91% (10/11)**	**47% (7/15)**

This leads to the first example having a visibility index of **91%** versus the second example having one of **47%**. Although on the surface such a simple scoring is appealing it does hide the underlying complexities of what is going on. For instance, in both examples, the opening of a file or application is initiated through the double-clicking of the mouse button. This effectively renders this function as invisible. In other words, it requires prior knowledge to do it. Therefore, an application can have a high visibility index but still have major issues around key functions that have a severe impact on usability. It is possible to weight the functions in different ways, such as the frequency of use or the impact on the user not knowing it is there. However, this is potentially highly subjective without real-world usage data.

It also does not handle the issue of prior knowledge, for example how many users will know to double-click the file or application icon.

However, this approach does several things. Firstly, it highlights the number of functions that have visibility issues that can then be reviewed by usability practitioners or the wider project team. Secondly, it shows the tradeoff between making functions visible and the visual clutter that would result if all functions were made visible. Finally, it makes comparison across potential designs or competing products possible. In this case, a traffic-light-style rating system is used to help navigate the scores. This leads to the table being a simple 'heat-map' and the style of this approach will be returned to later in the UEM development work.

Overall, despite the limitations the simple visibility index helps address the second initial research question:

> **IRQ 2:** *What problems is it causing users?*

This requires an understanding of:

> **SRQ 2a:** *What problems can it cause?*
> **SRQ 2b:** *Are these problems significant?*

In line with this and SRQ 2b, in particular, it was decided to look at a smartphone in more detail to explore further interfaces that are in widespread use.

4.2 Detailed Example

The device chosen for more detailed analysis was an Apple iPhone™ smartphone. The Apple mobile and tablet operating system, iOS™ was chosen because it is widely used (Apple Inc., 2016a) and mature, being in its tenth version (Apple Inc., 2018). The device chosen was an iPhone 7™ (see figure 4-4), as the latest in a series of iPhones, which at the time of testing over a billion had been made of different models (Apple Inc., 2016a) and subsequently by 2018 they reached the 2 billion mark for devices based on the same operating system (Apple Inc., 2018). The critical task of app launching and management was chosen as the target area for analysis. This is not only a significant task area but also provides a comparison to the previous work to highlight the evolution and differences across different types of interface (WIMP versus touch). This functionality is performed via home screens (see Figures 4-5 and 4-6) and it should be noted that for this version Apple calls each subsequent 'screen' of apps an *"additional home screen"* (see Figure 4-6), hence the use of the plural term (Apple Inc., 2016b). Out of the box, there are two home screens. In addition to the ability to launch and manage apps, the home screens provide shortcuts to various actions, for example, turning the torch on or launching the camera. The iPhone 7™ also has '3D Touch' interaction that enables users to access even more shortcuts by pressing the screen with higher force.

The configuration was based on the 'out of box' set of applications with the device configured as part of the set-up process. This initial starting configuration will vary slightly from user to user but will not make a substantive difference to the common functions available. Although in reality a user would download apps that they want and reconfigure their layout it seems an acceptable compromise to create a benchmark configuration based on the 'out of box' app set. The task sequence explores all the functions available starting from the first home screen with the device awake and unlocked.

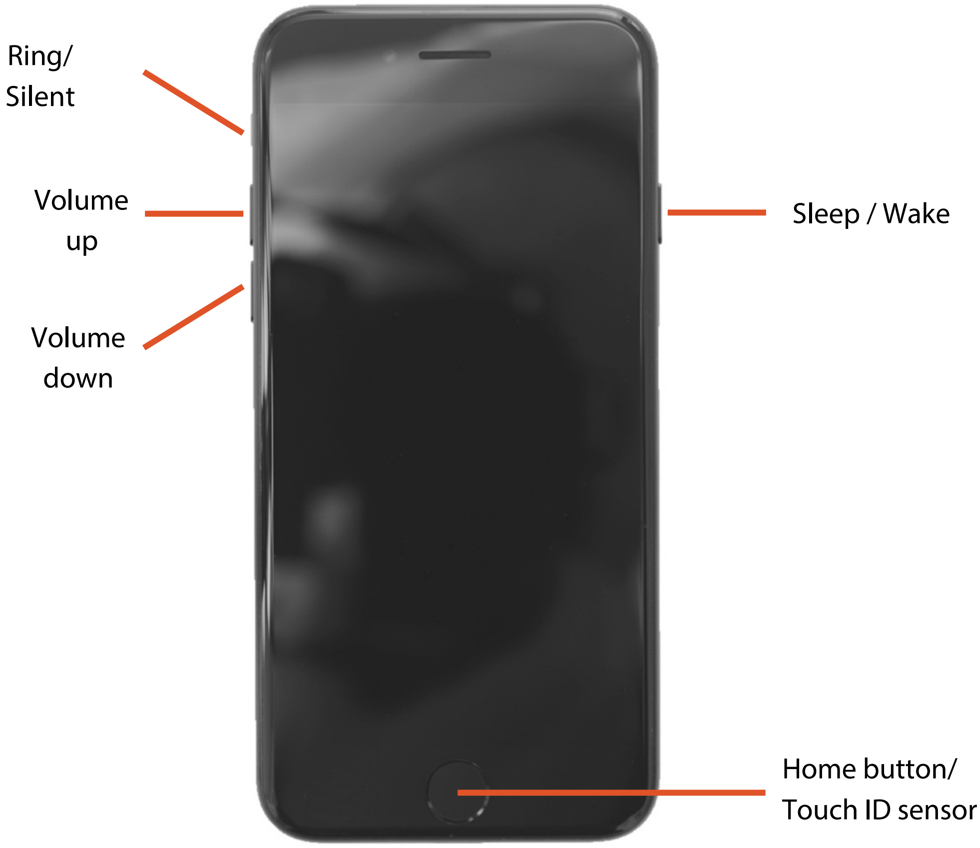


Figure 4-4: iPhone 7™ showing location of buttons

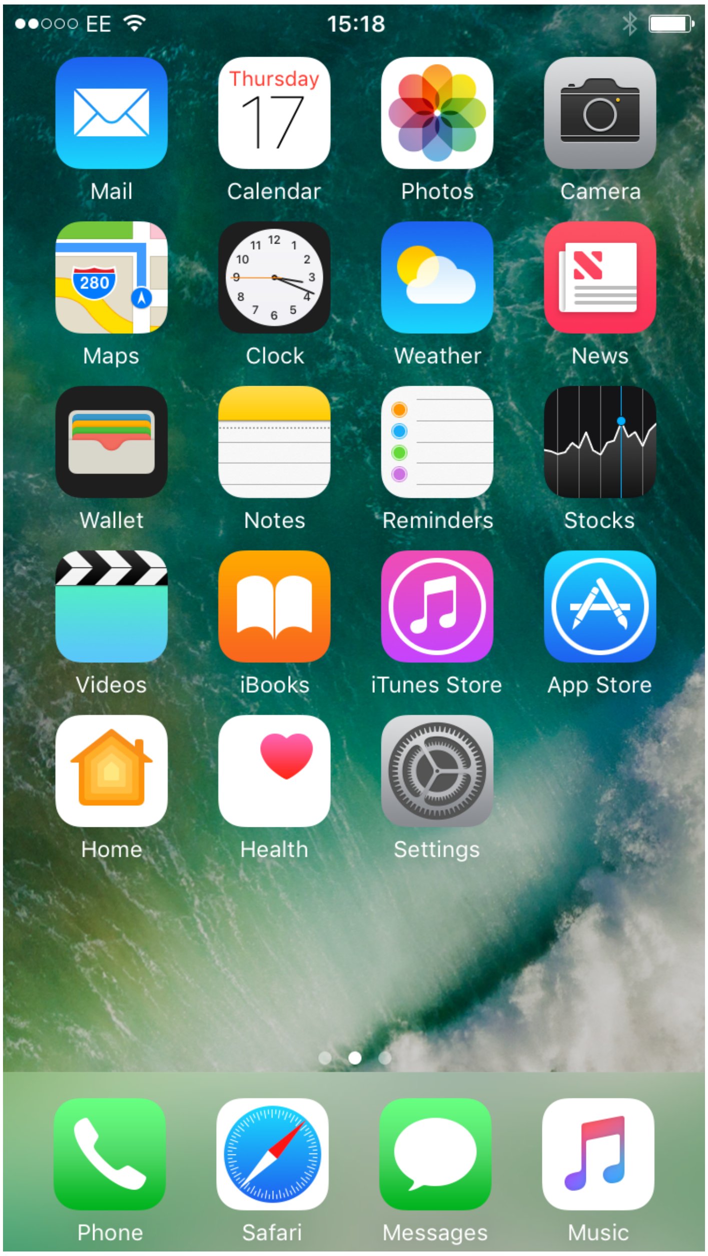


Figure 4-5: iOS 10™ home screen (Apple Inc., 2016b)

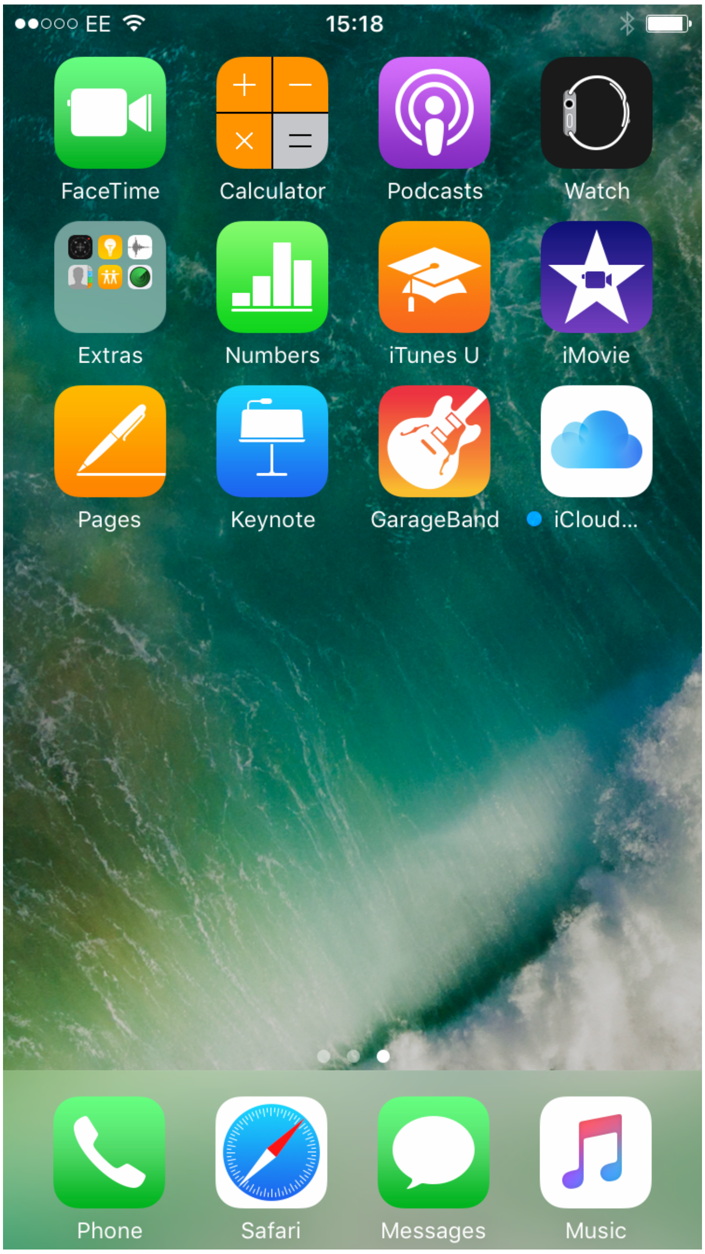


Figure 4-6: iOS 10™ additional home screen (Apple Inc., 2016b)

Interaction is broken down into the different modalities as follows:

- Buttons
- Touch
- Movement of the device
- Voice

Several aspects require definitions:

- **State** is the configuration of the phone at a point in time.
- **Initial state** is the starting point for the analysis. In this case, the phone is awake, unlocked and on the first home screen.
- **Level** refers to the number of states from the initial state that the user navigates through. It is equivalent to the levels in a traditional menu system. Accordingly, the initial state represents the top-level or level one.
- **Function** is defined broadly as an action presented to the user, that either presents a further range of actions or is an action in its own right, such as opening an app or a direct shortcut to an action within an app (e.g. creating a new message). To use the vernacular, a function is "a thing that a user can do". Such a broad definition is used to see how many of the "things a user can do" are visible. This definition is in contrast to a more rigorous one, such as Gero's (1990) function-behaviour-state, which would have potentially over-complicated an already detailed analysis.
- **Unique** refers to whether the function only appears once within the functions that are considered, in this case, functions available from the home screens. Therefore, the function may possibly appear elsewhere within iOS™ and the apps that it supports.
- **Visible** is defined as any graphical element that is there to indicate the presence of a function.

Where appropriate the terminology, describing the user interface, is based on the iPhone User Guide (Apple Inc., 2016b), however, terms are changed where greater clarity is required. For example, the difference between a light touch and a more forceful one on the home button is not distinguished completely in the Guide. For this case, a double light press is described as a 'touch' of the home button which activates the reachability mode as opposed to a double press with a greater force which is described as 'press', which activates the application switcher.

The analysis was recorded in a spreadsheet showing the available functions and the level they are at (Hosking and Clarkson, 2017). This is logically equivalent to a standard WIMP menu hierarchy. A handful of functions are available at all levels and this is noted accordingly. In practice, the process was iterative to find the best structure and grouping within the spreadsheet. This was done in conjunction with 'live' testing on an iPhone which also revealed further functionality appropriate for analysis.

The analysis should be seen as comprehensive but not exhaustive. This is due in part to the fact any specific configuration does not allow for all options to be available. Also, a full state transition diagram covering every operational situation (e.g. receiving a call when navigating the home screen or the variation in quick action menus depending on previous app usage) would add significant complication. However, for this analysis it is was deemed that the analysis was sufficiently comprehensive to highlight the issues around visibility.

4.3 Results

The approach resulted in 622 functions being analysed. As such they are too numerous to display them within the chapter but the full spreadsheet has been made available (Hosking, 2017). A snapshot of an example is shown in Table 4-3. This is the analysis of the Mail app. It shows the functions available, which are at level one and two in the hierarchy of functions. The action required to activate the function is recorded, as well as a description of the function. The final 3 columns then record: if the line item is a function, to enable the total number to be easily determined; if the function is unique within the set of functions being considered; and if the function is visible at the top level. These columns are then used to count the respective numbers. It should be noted that a function may become visible at a lower level but in terms of discoverability at the top-level, it is invisible.

Table 4-3: An extract from the analysis spreadsheet (Hosking, 2017)

	A	B	C	D	E	F	G	H	I	J
1		Touch Screen				Action	Function Description	Function	Unique	Visible
3		L1	L2	L3	L4					
108		Mail								
109		App icon				Press	Launch App	Y	Y	Y
110		App icon				Touch & hold →	App Configuration	Y	N	N
111		→	Configure App			Press 'cross'	Delete App	Y	N	N
112		→	Configure App			Drag to new location	Move App	Y	N	N
113		→	Configure App			Drag onto another app	Create Folder	Y	N	N
114		→	Configure App			Drag onto a folder	Add App to Folder	Y	N	N
115		→	Configure App			Drag to dock	Add to dock	Y	N	N
116		→	Configure App			Drag right	Create New Home Screen	Y	N	N
117		App icon				3D press →	Quick Actions Menu	Y	N	N
118		→	Quick Action			Press	New Message	Y	Y	N
119		→	Quick Action			Press	Search	Y	Y	N
120		→	Quick Action			Press	All VIP emails	Y	Y	N
121		→	Quick Action			Press	All inboxes folder	Y	Y	N
122		→	Quick Action			Press	VIP 1 emails	Y	Y	N
123		→	Quick Action			Press	VIP 2 emails	Y	Y	N
124		→	Quick Action			Press	VIP 3 emails	Y	Y	N
125		→	Quick Action			Press	VIP 4 emails	Y	Y	N

Table 4-4 provides a numerical summary of the entire analysis for the 4 different modalities considered. A percentage of the functions that are visible is calculated. Also, a percentage of the visibility of unique functions is calculated separately. This is done because it considers each repeated function, e.g. deleting an app, as a single instance on the basis that if you cannot see app deletion for one app you cannot see it for all the apps. Conversely, if you do 'know' it is there it reduces the impact of the lack of visibility of for all instances across an application. The combination of the two numbers helps give a more complete picture, particularly in the case where there is a lot of functionality that is repeated in different parts of an interface.

Table 4-4: A numerical summary of the functions analysed and their visibility

Modality	Number of Functions	Number that are Unique	Number that are Visible	% that are visible	% of unique that are visible
Buttons	23	17	5	22%	29%
Touch	597	266	43	7%	16%
Movement	1	1	0	0%	0%
Microphone	1	1	0	0%	0%
Total	**622**	**285**	**48**	**8%**	**17%**

4.4 Discussion

The bare numbers present what appears to be a stark situation with regards to the high number of functions (622) and the low number of these that are visible (8%) even if repeated functions are removed the visibility figure only rises to 17%. The situation, concerning the total number of functions, would be increased further as users download additional apps. If one was also to include all the in-app functions, then clearly the number would increase dramatically.

However, the situation is far more nuanced than the bare numbers indicate. The general issues around this were highlighted earlier and apply in this example as follows:

- The relevance of each function is not weighted in any way, for example with their importance or frequency-of-use. As it stands the home screens appropriately prioritise the frequent and important task of launching apps. Secondary functions such as moving or deleting an app are made 'visible' by the 'touch and hold' of the app. This has the big advantage of reducing visual clutter. Indeed, it would be unrealistic to make all the functions visible.

- The numerical analysis does not consider prior knowledge of the users. In other words, their experience of using gestures on a touch device.
- There is a commercial imperative of offering new features to maintain sales. The needs of different users need to be balanced, in other words, there is a 'function' versus 'complexity' trade-off.
- It only covers the 'missing' and not the 'missed' or 'misunderstood' of the 3 M's Model.
- Finally, the score is not correlated to usability. This would require testing with users and something that should be considered for future work.

It can be argued that iOS™ strikes an appropriate balance between offering a rich set of functions and reducing visual clutter with a focus on the high priority and high-frequency functions. However, the Microsoft Office™ example cited earlier is a cautionary tale regarding users not finding functionality and there is a real concern for novice or older users in particular. The numerical analysis should not be used as a simple 'good to bad' scoring system but instead it should be used to highlight the overall status of the visibility of a system and to review whether the trade-offs that have been made are appropriate. In addition, support for novice or older users can be looked at, for example providing a simple mode or on-screen prompts that can be toggled on and off.

4.5 Conclusion

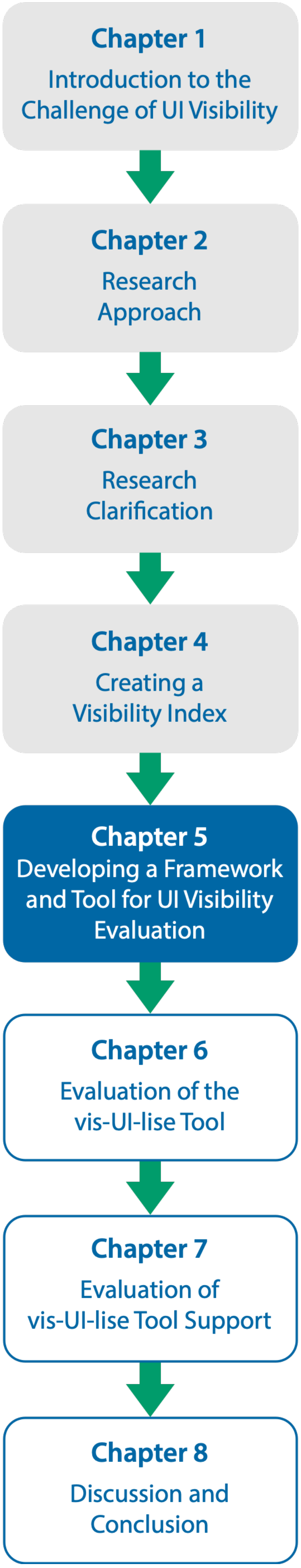
This visibility index gives credence to the "usability crisis" (Norman and Nielsen, 2010). In particular, it points to current touch interfaces in smartphones and other devices being a retrograde step with regards to UI visibility. In a sense, they represent a return to the visibility problems of command-line interfaces described in the previous chapter. The 'remember and swipe' of a touch interface being analogous to the 'remember and type' of a command line.

The index also addresses the second research question.

> **IRQ 2:** *What problems is it causing users?*
>
> This requires an understanding of:
>
> **SRQ 2a:** *What problems can it cause?*
> **SRQ 2b:** *Are these problems significant?*

It highlights the magnitude of the 'raw' problem in a widely used device and interface style. However, the index needs to be used with appropriate caution. It has significant headline appeal i.e. claiming that only 8% of 622 functions are visible. Yet there are many contextual factors outlined in the previous section that need to be considered. For this research, it serves its purpose in demonstrating the significance of the problem and helping create a baseline understanding. It also opens up the possibility of another research thread that could be explored. In particular, looking at a range of devices and trying to correlate the index with usability measures. This would provide further contextual understanding and the potential use of the index as a proxy measure for usability.

The index does not fully address the gap identified in the literature review regarding the lack of consideration of visibility in current UEMs. In particular, it only covers the 'missing' element of the 3 M's Model and not problems with functions being 'missed' or 'misunderstood'. The next chapter builds on this capability to identify the 'missing' and extends it to address all three M's.


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 5

Developing a Framework and Tool for UI Visibility Evaluation

5.1 Introduction[1]

The previous Chapter demonstrated an approach for quantifying the proportion of functions that are effectively 'missing' from a visibility perspective through a visibility index. This does not address the issue of functions that are 'missed' or 'misunderstand'. This chapter addresses this through the third initial research ruestion (IRQ) and more specifically through the two supplementary research questions (SRQs) as follows:

IRQ3: *What can be done to improve it?*

> **SRQ 3a**: *How does a cognitive-based framework for understanding and representing the visibility of user interfaces compare to one based on a simple visibility threshold?*
>
> **SRQ 3b***: Can such a framework be embodied in a tool that predicts more usability problems than current approaches?*

These questions are addressed through the development of a high-level visibility framework and detailed interaction model (SRQ 3a) which in turn leads to the creation of a usability evaluation tool (SRQ 3b).

From the DRM perspective, the visibility index represents the culmination of the Descriptive Study 1 and the shift in the work to developing an intervention marks the start of the Prescriptive Study. However, the high-level visibility framework and interaction model described in the chapter in many ways acts as a bridge between these two stages.

5.2 Method

The interaction between the descriptive (models) and prescriptive (tool) elements of the DRM are highlighted in Figure 5-1. This shows the core

---

[1] This chapter draws extensively on the published work of Hosking and Clarkson (2017; 2018a; b) with significant additions and elaboration.

elements that helped develop the models and associated tool. In practice, it was highly iterative, not only in terms of the contributory areas of literature and product evaluation but also the conceptual elements of the model and its presentation in a form that would be accessible to usability practitioners. This interaction is shown by the various arrows in the diagram.

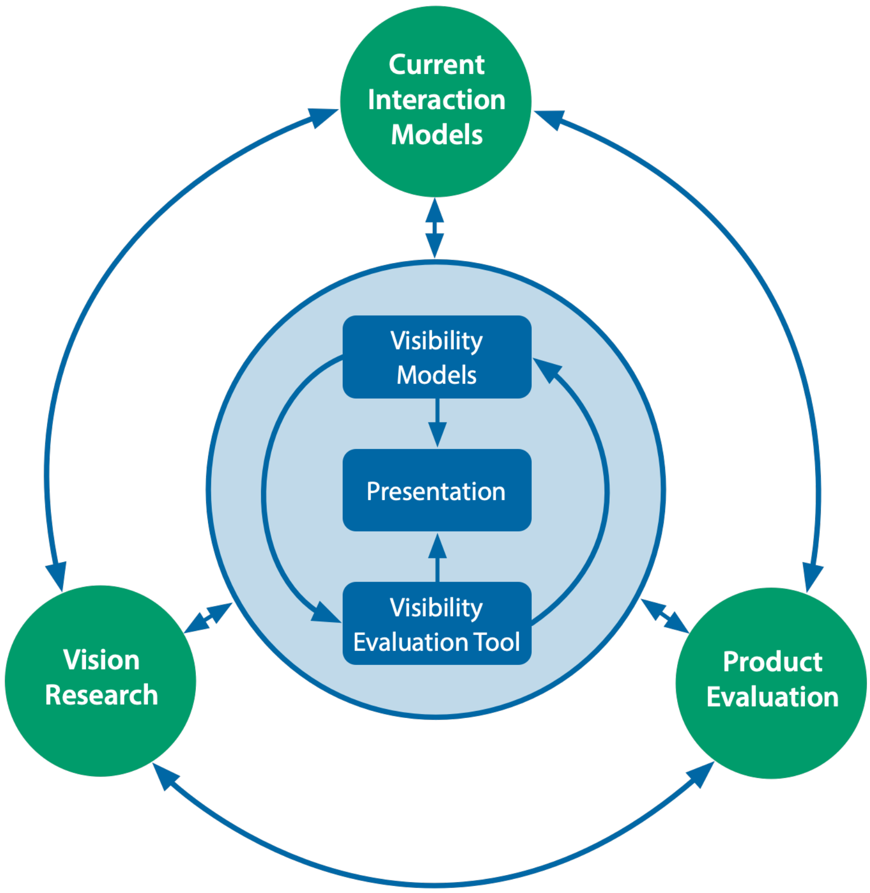


Figure 5-1: Overview of the contributory areas in the development of the models and tool

It is 'models' plural because of the development of a high-level visibility framework and a detailed model of interaction that elaborates on this. This work builds on the literature review, in particular, the exploration of different interaction models. This analysis is described in the following section.

5.3 Interaction Models Analysis

Table 5-1 below is a simplified version of Table 3-6 from Chapter 3. This formed the theoretical basis for defining the high-level framework and detailed interaction model for UI visibility.

Table 5-1: A summary of different interaction models (simplified from Table 3-6)

Model	Description
Perceptual Cycle (Neisser, 1976)	A generic, seminal cognitive model of how humans interact with the world
Model Human Processor (MHP) (Card, Moran and Newell, 1983b)	A simplified cognitive architecture
GOMS Family of Models (Card, Moran and Newell, 1983a; John and Kieras, 1996)	A simplified cognitive architecture for predicting performance related to MHP
7 Stages of Action (Norman, 1986)	Model of action that includes the gulfs of execution and evaluation
User Action Framework (Andre et al., 2001)	A simplified form of Norman's (1986) 7 stages of action
Task-Artifact Cycle (Carroll, Kellogg and Rosson, 1991)	A simple model showing the relationship between the task and the device
Object-Action Interface Model (Shneiderman and Plaisant, 2005, pp.95-101)	Relates the task to the interface in the context of direct manipulation
Cognitive Dimensions (Green, 1989)	A set of heuristics for visual design
Context of Use (Bevan, 1995)	A high-level model of the context of use to aid the structuring of measures for quality of use
ACT-R (Anderson and Lebiere, 1998)	Sophisticated cognitive architecture
Three-stage Model of Vision (Ware, 2003)	A Simplified three-stage model of vision
Capability-Demand Model of Interaction (Persad, Langdon and Clarkson, 2007; Waller, Langdon and Clarkson, 2009)	Describes how exclusion can occur when the demands of a product exceed a user's capabilities
Physics, Physiology, Psychology Framework (Green, 2008)	A broad framework for vision including perceptual exploration
Human Information Processing Stages (Wickens et al., 2013)	Model showing process stages relevant to task execution
PCA [Perception-Cognition-Action] Model (CENELEC, 2015, p.24)	The model adopted by standards bodies for medical device development and evaluation

A number of key elements from the models were identified and influenced the development process. These are summarised in Table 5-2.

Table 5-2: Summary of the key relevant properties & example papers that embody them

Model Property	Key Example
A staged process	7 stages of action (Norman, 1986)
A cyclical process	Perceptual cycle (Neisser, 1976)
Top-down & bottom-up processing	Human information processing stages (Wickens et al., 2013)
Exploratory process of understanding (perceptual exploration)	Physics, physiology, psychology framework (Green, 2008)
Progressive building of visual understanding (sensation – attention – perception)	Three-stage model of vision (Ware, 2003)
A series of hurdles to be overcome	Capability-demand model (Waller, Langdon and Clarkson, 2009)
Context dependence	Context of use (Bevan, 1995)

Box (1979, p.202) stated *"all models are wrong, but some are useful"* and in his earlier work (1976, p.792) he argues for *"economical description of natural phenomena"* and the avoidance of *"over-elaboration"*. This is clearly a laudable aim but made harder in this case by the representational tensions, such as portraying a staged process versus one that emphasises top-down and bottom-up processing. Box and Draper (1987, p.74) unpacked this issue by saying *"…the practical question is how wrong do they have to be to not be useful"*. Trade-offs were guided by having a focus on addressing the 3 M's and the visibility problems found from the product analysis. In addition, consideration was given to making it comprehensible to the primary target audience of usability practitioners.

5.4 High-level Framework

The development of the high-level model was inspired by Green et al. (2008) with their layers of physics, physiology and psychology. These categories neatly cover the nature of light and its environmental context (Physics); the sensing of this light by the eye (Physiology); and critically the cognitive element (Psychology). However, what it lacks is the interactive element that is covered well by models such as Norman's (Norman, 1986) stages of action. Figure 5-2 shows how interactivity is added through the introduction of 'product' which is broken down into the 'program' providing the underlying functionality and 'presentation' representing the user interface.

The issue of 'missing' is driven by the fundamental design of the 'program'. Things that are visible and 'missed' can result from a breakdown in: 'presentation'; the physical environment (physics); the user's eye (physiology); to the focus of their attention (psychology). Finally, 'misunderstanding' comes primarily from the interpretation (psychology) of the 'presentation'. This model is high-level and frames the overall area of UI visibility and is thus called a framework guiding the development of the detailed model.

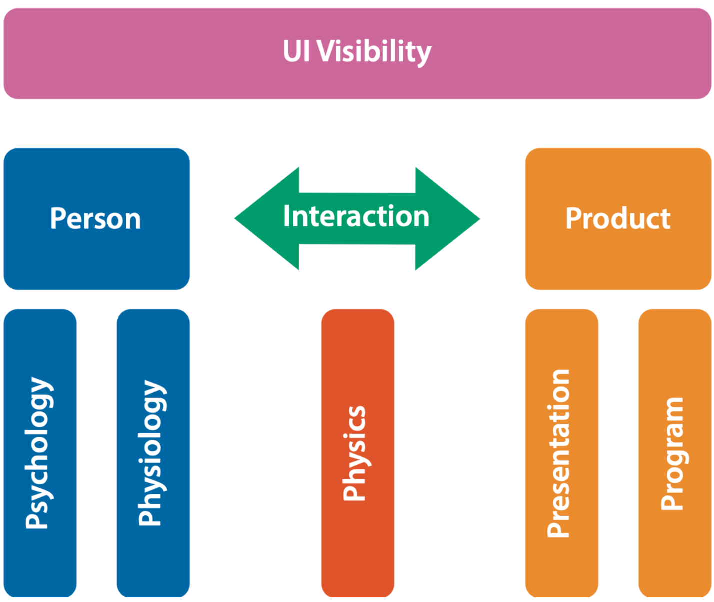


Figure 5-2: High-level framework of UI visibility [from Hosking and Clarkson (2018a)]

5.5 Detailed Model and Tool Development Requirements

Having established a high-level framework, it is prudent to consider the requirements for the detailed model and associated UEM tool. This helps provide the focus that ensures the model is useful (albeit 'wrong') and provides the foundations for a UEM tool. Gray and Salzman's (1998) robust critique of five well established UEMs is a sobering benchmark for anyone attempting to develop a new UEM. A response to this in terms of guidance on development is provided by Blandford and Green (2008) and in a slightly different form by Blandford et. al (2008). A comparison of the differences is shown in Table 5-3 below.

Table 5-3: Comparison of UEM development requirements

Blandford & Green (2008)	Blandford et. al (2008)
Validity	Internal Validity
Scope	External Validity
Reliability	Productivity
Productivity	Practicalities
Usability	Persuasive Power
Learnability	Analyst Activities
Insights Derived	Scope

These two lists are synthesized as follows to provide a high-level requirements list for the development. Items in square brackets relate the rewording of certain items from the original descriptions which are shown in Table 5-3.

1. **Scope & Positioning** [Practicalities]: What the proposed UEM will cover and how it is positioned concerning the development process and other evaluation approaches.
2. **Output** [Insights derived]: The specifics of the output e.g. quantitative and qualitative aspects.

3. **Validity:**
    a. **Internal** [Reliability]: Internal validity concerns the consistency across different evaluators (inter-rater reliability). Hertznum and Jacobsen (2001) highlight the significant problems of the 'evaluator effect' in terms of the variation of output across different evaluators.
    b. **External:** External validity addresses how well any results relate to 'real-world' performance.
4. **Tool Usability:** Several requirements have been grouped under usability in line with the definitions of usability discussed in Chapter 3 (see Figure 3-24).
    a. Productivity (effectiveness & efficiency)
    b. Learnability
5. **Persuasive Power:** Ultimately evaluation is only of value if it can effect change. Therefore, the approach and its output's ability to persuade designers to make appropriate changes is critical.

When applied to this research it leads to the following initial outline description of the UEM as follows:

Scope & Positioning: The aim of the UEM tool is threefold as follows:

a) Identify visibility issues covering 'missing', 'missed' and 'misunderstood' functions
b) Rating of the severity of the problems
c) Predicting what usability problems these may cause

As such it is primarily positioned as a tool suited to early-stage evaluation, in advance of user testing either to address issues before testing or to highlight areas of concern to focus user testing on. The aim is not to replace existing UEM's but to complement them by providing greater depth of insight around visibility issues. The primary target users are usability practitioners.

Output: With the threefold scope, outlined on the previous page, there are distinct outputs for each area as follows:

a) Screenshots highlighting the problems
b) A simple rating scale of the severity of the visibility problems
c) A set of usability problem predictions linked to the visibility problems with a prediction of the likelihood of the problem occurring and its potential impact

With regards to the requirements around tool usability and persuasive power these were further expanded into the following requirements:

R1 The starting point was to use **standard cross-platform 'office' tools** (word processor, spreadsheet, presentation) as they are widely available and familiar to many people.

R2 The aim to was to have a **flexible template** that would work for different product UI form factors e.g. small and large displays.

R3 In addition to flexibility around different form factors, the aim was for the tool to be **easily extensible by the evaluator** to cope with different project needs.

R4 **Recording and rating problems quickly** were deemed to be essential to reduce evaluation workload and any reluctance of practitioners to adopt a new tool.

R5 With 'persuasive power', the aim was to have **a form that could work for presenting to stakeholders** and as a standalone report for individual practitioners to review.

R6 Again, in relation to 'persuasive power', the presentation must **allow key problem areas to be found quickly** on the basis that there may be a large number of issues to get through when reviewing with stakeholders.

5.6 Detailed Interaction Model & Associated UEM Tool

An initial detailed model was developed, influenced by the interaction models in Table 5-2. ‘Physiology’ is described in terms of ‘sensation’ and breaks ‘psychology’ into ‘attention’ and ‘perception’ (Green, 2008; Wickens et al., 2013). The functions that the ‘program’ provides and ‘presents’ are broken down based on the interactive cycles of Norman (1986) and Andre et al. (2001) as follows:

1. Can the user 'see' the **function** (interface control)
2. Can the user 'see' how the function (interface control) **operates**
3. Can the user 'see' **feedback** that the function (interface control) has been operated correctly

The distinction between the visibility of the function and the operation of the function is important as the user may be able to easily work out what a control does but not know how to operate it or vice versa. This is particularly an issue with multi-finger, multi-gesture touch interfaces that have no or limited visibility with regard to the use of such gestures. This control-level view, shown in the bottom half of Figure 5-3, was allied to higher-order cognitive processes related to goals and prior knowledge. This was derived from Wickens et al. (2013) top-down and bottom-up processing model.

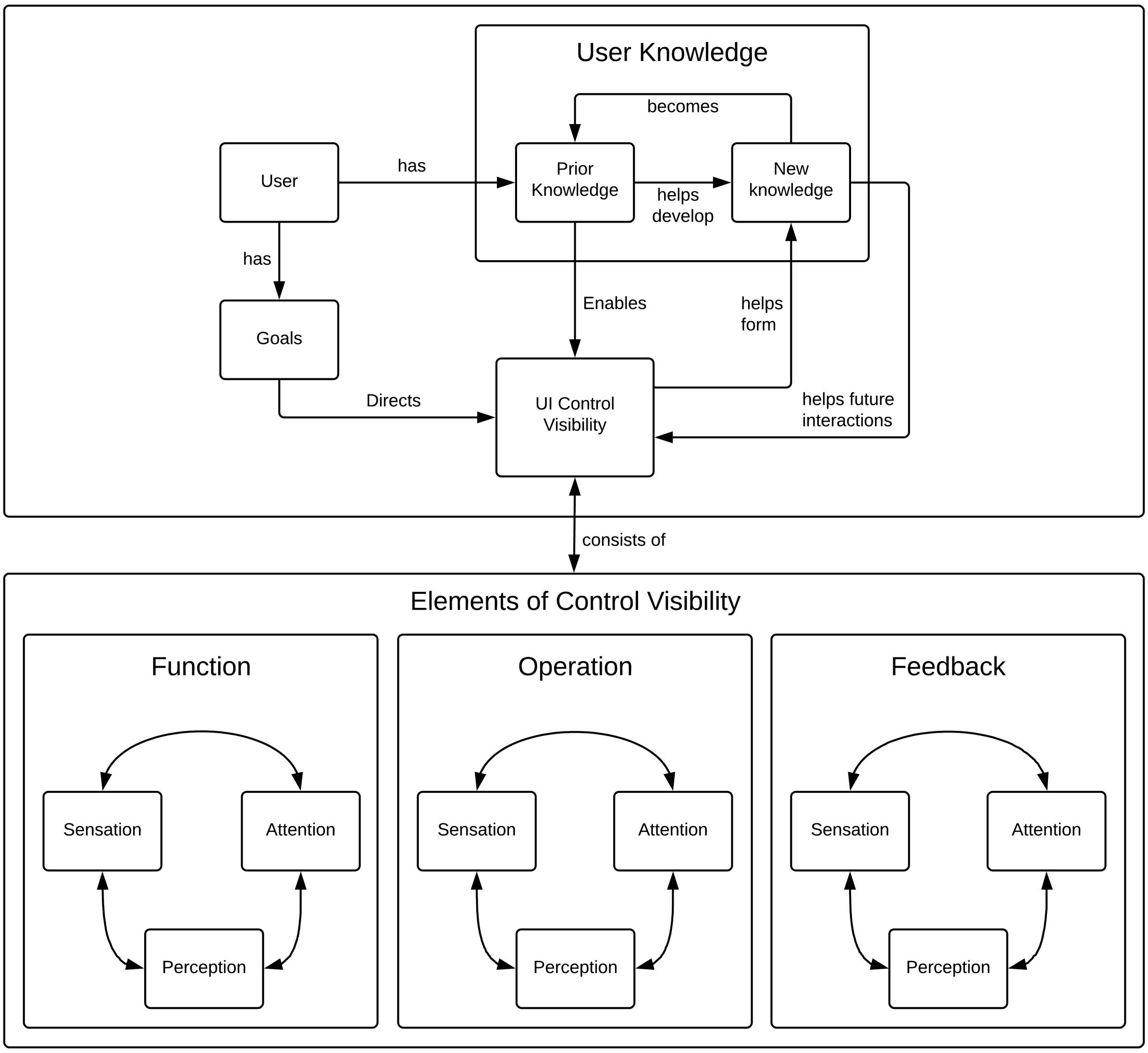


Figure 5-3: First detailed model of user interface visibility

Although the model brings together many of the key elements it suffers the *"over-elaboration"* that Box (1976) describes and came primarily from a theoretical standpoint. This was reviewed in light of the specific problems found with various products in the Descriptive Stage 1. It was decided to compromise by focussing on a simpler staged approach and drawing on the capability-demand model of Persad, Langdon and Clarkson (2007). This was combined with the notion of the stages being a series of 'hurdles' that the user has to get through (Nicholl, 2017). This led to the revised model shown in Figure 5-4.

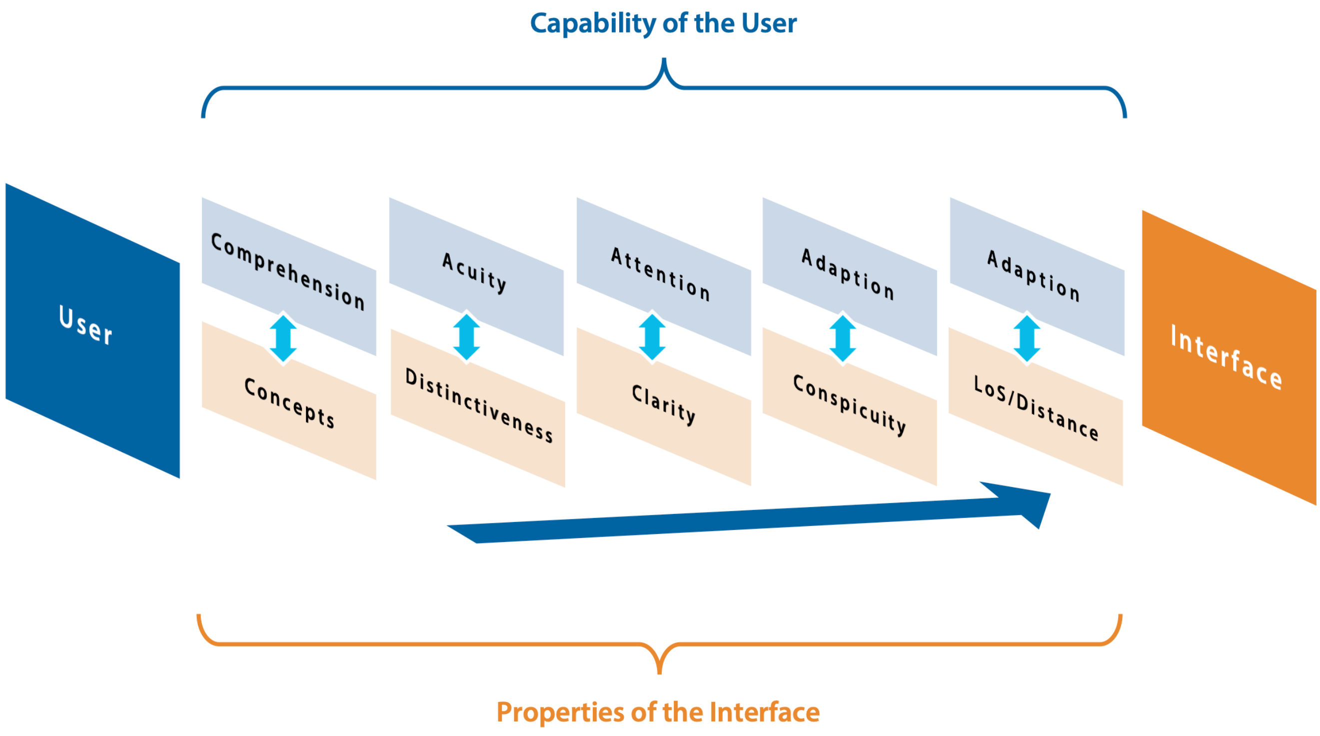


Figure 5-4: Second detailed model based on a demand-capability approach (LoS stands for Line of Sight)

A comparison of the two models is shown in Table 5-4. This shows the transition in the second model to a more performance-oriented language e.g. sensation becomes acuity. Two additional elements are added. 'Differentiation' concerns the ability to distinguish between functions that have a similar visual appearance. This problem was identified in the product examples often due to the proliferation of icons on many interfaces. Finally, 'adaption' was added to cover the ability of the user to achieve line-of-sight with the visual elements and focus on them.

Table 5-4: Comparison of first and second candidate detailed models

First Model	Second Model	
	User Capability	Product Demand
Sensation	Acuity	Clarity
Attention	Attention	Conspicuity
Perception	Comprehension	Concepts
	Differentiation	Distinctiveness
	Adaption (Focus)	Line of sight/distance (Location)

The hurdles are strongly linked to the literature on vision (see Section 3.5 in Chapter 3). In that section vision was defined as follows:

> *"It is an edge-detecting, dynamic, very slightly delayed, selective, blank-filling, prior-experience-combining, object-inferring, distance, direction and speed-estimating, action-oriented system. It is part of a prolific inference engine, making sense out of an incomplete, noisy, sensory input. Vision is an attentionally-driven, zooming spotlight that outputs a 3D, colour, stabilised, immersive representation of the world that enables effective action within it."*

The theoretical understanding, in combination with an exploration of visibility issues in different products, and through various iterations, helped define the hurdles as follows:

Capability: **Focus** – Demand: **Location**

The language of this was modified to make it more practitioner-friendly. The original use of the word 'adaption' (now 'focus') is technically precise but not necessarily familiar to practitioners. As was discussed in the literature review the ability to 'adapt' and focus on items that are near declines appreciably with age and impacts over 1 billion people worldwide (Wolffsohn and Davies, 2018). Therefore, highlighting this was deemed to be a key component for the visibility model. The car is perhaps one of the best examples of the demand on our visual accommodation (focus) of differing locations, from looking at the road ahead to items on the centre of the dashboard, then nearer with the instruments behind the steering wheel and closer still with buttons on the steering wheel. Also, controls may be obscured by the steering wheel.

Capability: **Acuity** – Demand: **Clarity**

Having established how well a user can focus on the necessary elements of the user interface the next issue that arises is the fundamental clarity of the controls. In other words, to understand whether the key visual elements are in the 'window of visibility' (Watson, Ahumada and Farrell, 1986; Watson, 2009; Watson and Ahumanda, 2016). Breaking down the visual attributes and considering them in terms of their contrast and size is the key to understanding clarity.

Capability: **Attention** – Demand: **Conspicuity**

However clear visual elements are, they rarely exist in isolation and have to compete for our attention amongst other distractions, including other UI controls. Will the attributes of the 'target' user interface elements stand out, in other words, to be conspicuous, from their 'distractors'? This is the heart of understanding whether our attention will be appropriately guided at the right time (Wolfe and Horowitz, 2017).

Capability: **Differentiation** – Demand: **Distinctiveness**

Our attention may be appropriately guided, but can we differentiate one potential control from another? Figure 5-6, which is discussed later on, shows how difficult this can be and moves to higher-order cognition in determining what a control might be.

Capability: **Comprehension** – Demand: **Concepts**

Finally, if the other hurdles have been successfully cleared can the user comprehend what a control is for (function), how it works (operation) and if it has worked correctly (feedback).

The critical advance in this model was to enable human-capability to be mapped to the properties of the product which ultimately is required to understand how 'good' the visibility of different aspects of a UI is. This leads to representing things just in terms of the demands of the product as shown in Figure 5-5. A specific product example of this is shown in Figure 5-6 for a car heating and ventilation system (HVAC). It represents the task of activating the rear window demisting (button highlighted with a yellow rectangle). The flow of the hurdles is from right to left, which may be confusing at first but is a compromise to show the 'zooming in' (LaBerge, 1983) on the function of interest. This example highlights well a particular instance of the problem of 'differentiation' as the button to the right is for front window demisting. The only difference is the shape of the 'window' of the icon on the button compared to the rear window demisting.

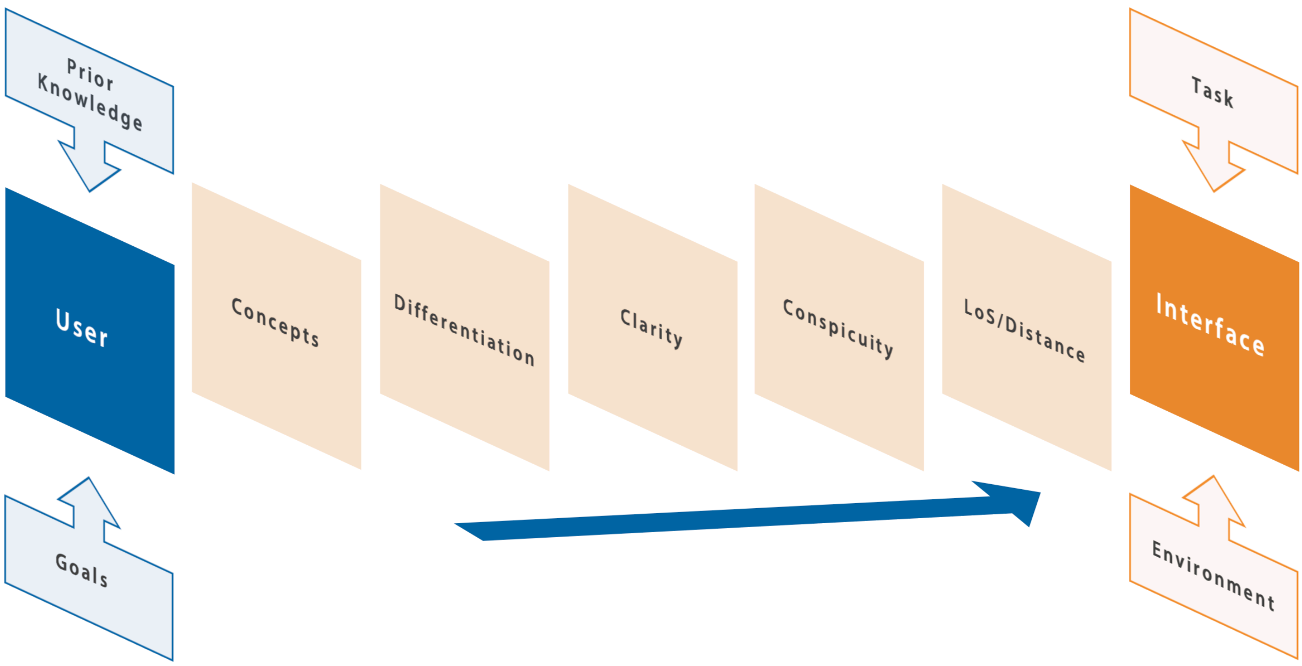


Figure 5-5: Variant of the detailed model that just shows the product properties

Figure 5-6: Car HVAC interface broken down into a series of hurdles for the task of activating the rear heated window (LoS stands for Line of Sight)

5.6.1 Hurdle questions

Having developed a detailed model that represents interaction as a series of hurdles, it is possible to structure their assessment by posing a range of questions for each hurdle. Thus, the model easily progresses into the beginnings of an evaluation tool. Before proceeding with this it is worth reflecting on whether this is the most appropriate way forward. An alternative would be to consider modifying or extending existing UEMs. Table 3-10 from Chapter 3 provides a summary of such UEMs which was used to demonstrate the gap in the explicit evaluation of UI visibility. Indeed, it is this gap that justifies a focus on UI visibility alone to avoid it being constrained by the structure of an existing method. Additionally, broadening the boundary of UI visibility to take a cognitive perspective means it would overlap with the cognitive elements of existing evaluation approaches. This would make it hard to delineate between which cognitive elements relate directly to vision as opposed to other areas of cognition. Indeed, it enables the boundaries of where UI visibility begins and ends with regard to the cognitive dimension to be explicitly explored. It is also reasonable that this dedicated focus increases the chances of successfully addressing SRQ 3b concerning whether a framework can be embodied in a tool that predicts more usability problems than current approaches. As was stated earlier in the chapter the aim is not to replace existing UEMs but to complement them by providing deeper insights into UI visibility issues. Therefore, this dedicated and standalone approach is warranted at this stage in the research. It does not stop the later integration with existing UEM tools once UI visibility has been adequately explored.

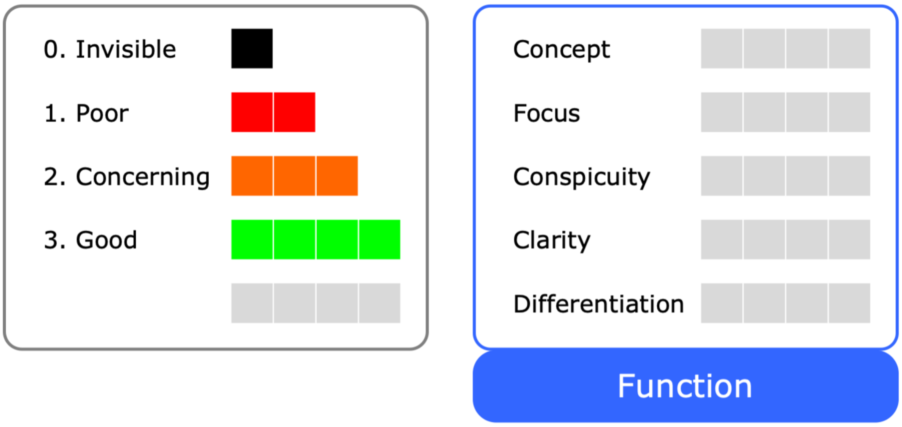
Returning to the hurdles and questions to evaluate them, a set of questions for this are detailed in Table 5-5. There is a 'main' question for each hurdle, with additional supplementary questions to further unpack the nature of the hurdle. This wording represents a refined version from the testing conducted with practitioners in Descriptive Stage Two and described in Chapter 7. In this latest version 'line of sight' and 'distance' has been replaced with the words 'focus' and 'location'. The first question is defining the 'concepts' related to the control/function that the visual attributes are trying to convey. This question is a critical step in the analysis and proved the most problematic when testing with practitioners (which we will return to in Chapter 7). In essence, if you do not know what needs to be conveyed then it is not possible to determine whether things are visible or not. For example, if the control is a button that will initiate a specific action, does it convey what the action is and that its operation is performed by pressing it?

Table 5-5: A series of 'hurdle' questions for the five hurdles in the detailed visibility model [modified from Hosking and Clarkson (2018b)]

Hurdle Question and their associated sub-questions
1. Are the concepts (metaphors) used to portray the 'function'/'operation'/'feedback'* of the control comprehensible to the user? 1.1 What are the concepts of the 'function'/'operation'/'feedback'?* 1.2 How are these concepts conveyed visually? 1.3 Are they familiar concepts to the user? (check against real examples) 1.4 How well are these concepts represented and are there elements missing? 1.5 Are there general variations of this concept that could cause confusion? 2. Is the location of the user interface control such that the user can focus on it? 2.1 Does the user have to move to get line-of-sight? 2.2 Is the distance such that the user can focus on it? 3. Is the user interface control sufficiently conspicuous that it grabs the user's attention at the appropriate time? 3.1 Is it in the central visual field? 3.2 Is it where the user would expect it to be? 3.3 How many other related controls are there? 3.4 Does it stands out against other controls/background? 4. Are the key visual parts of the user interface control of sufficient clarity (size and contrast) that the user can resolve them? (i.e. within the range of the user's visual acuity) 4.1 What are the distinguishing graphical features of the UI control? 4.2 What size are they? 4.3 What is the level of contrast compared to their background? 5. Is the user interface control sufficiently distinctive from other controls that the user can correctly differentiate them? 5.1 How different is it from other controls visible at the same time? 5.2 How different is it from other controls visible at other times? 5.3 Could it be confused with commonly used graphics/symbols that indicate something different? ** The hurdle questions are adapted and repeated accordingly for 'function', 'operation' and 'feedback'.*

5.6.2 Rating Scales

Section 5.5 outlined requirement R4 regarding the rapid rating and recording of visibility issues allied to R6 allowing the quick assimilation of visibility hotspots. The starting point for this was a simple traffic light style rating scale with the addition of a black rating to address functions that have no visible attributes ('missing'). The traffic lights are mapped to the words as shown in Figure 5-7. The rating scale gives a percentage probability of occurring for a user population. For example, 'poor' is likely to lead to problems in more than 10% of users, whereas 'good' in less than 1% of users. This will differ from 'first use' to 'repeated use', but provides some level of guidance. To put this in context even a 1% occurrence rate with a user base of millions can create substantial support problems. The rating scales are repeated for 'function', 'operation' and 'feedback' with a blue label under the scale showing which one it is.



0. Invisible	No visibility and is highly likely to cause usability issues
1. Poor	An issue likely to contribute to some level of usability problem
2. Concerning	An issue that may lead to usability problems
3. Good	Unlikely to directly cause usability problems
'AUTO' is put across the 4 cells when a function is automatically performed by the software/device	

Figure 5-7: A simple 4 level rating scale from 'invisible' to 'good'

Figure 5-8 shows an earlier candidate approach with the 5 hurdles compressed into 4 by combining 'focus' and 'clarity' into a single 'legibility' scale. A radar style plot was chosen to make it compact and to provide a strong visual appeal to try and address the 'persuasive power' requirement R6. This approach was used in the early evaluation of different devices (Figures 5-10, 5-11, 5-12 & 5-13) before switching to the final one shown in Figure 5-7 and 5-14. The final one also has the advantage that it is much easier to enter the values that the radar plot style format.

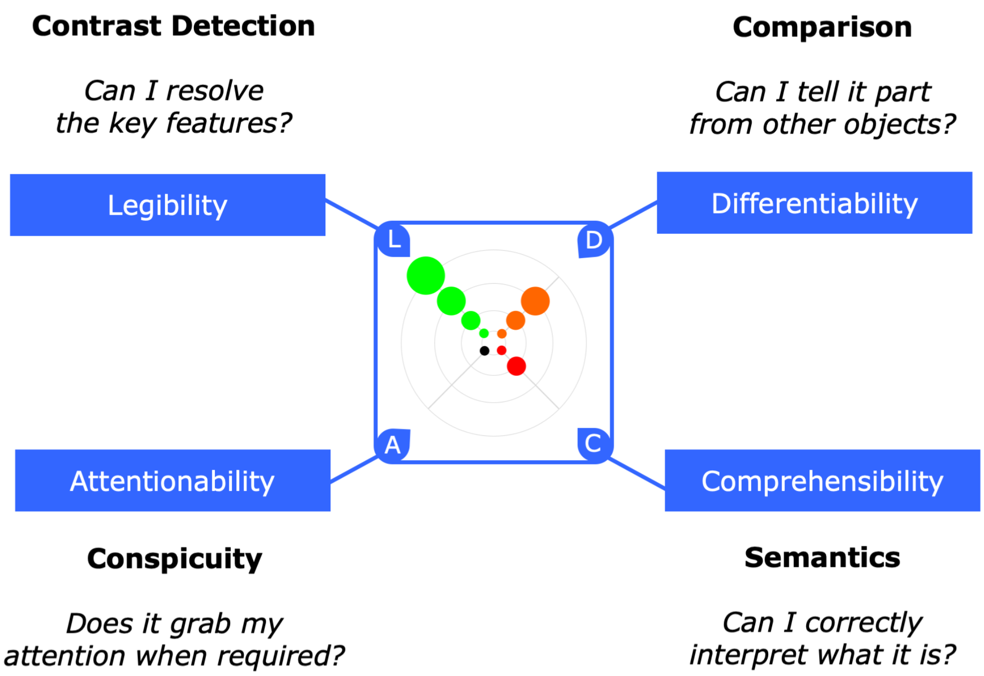


Figure 5-8: An early prototype rating scale display

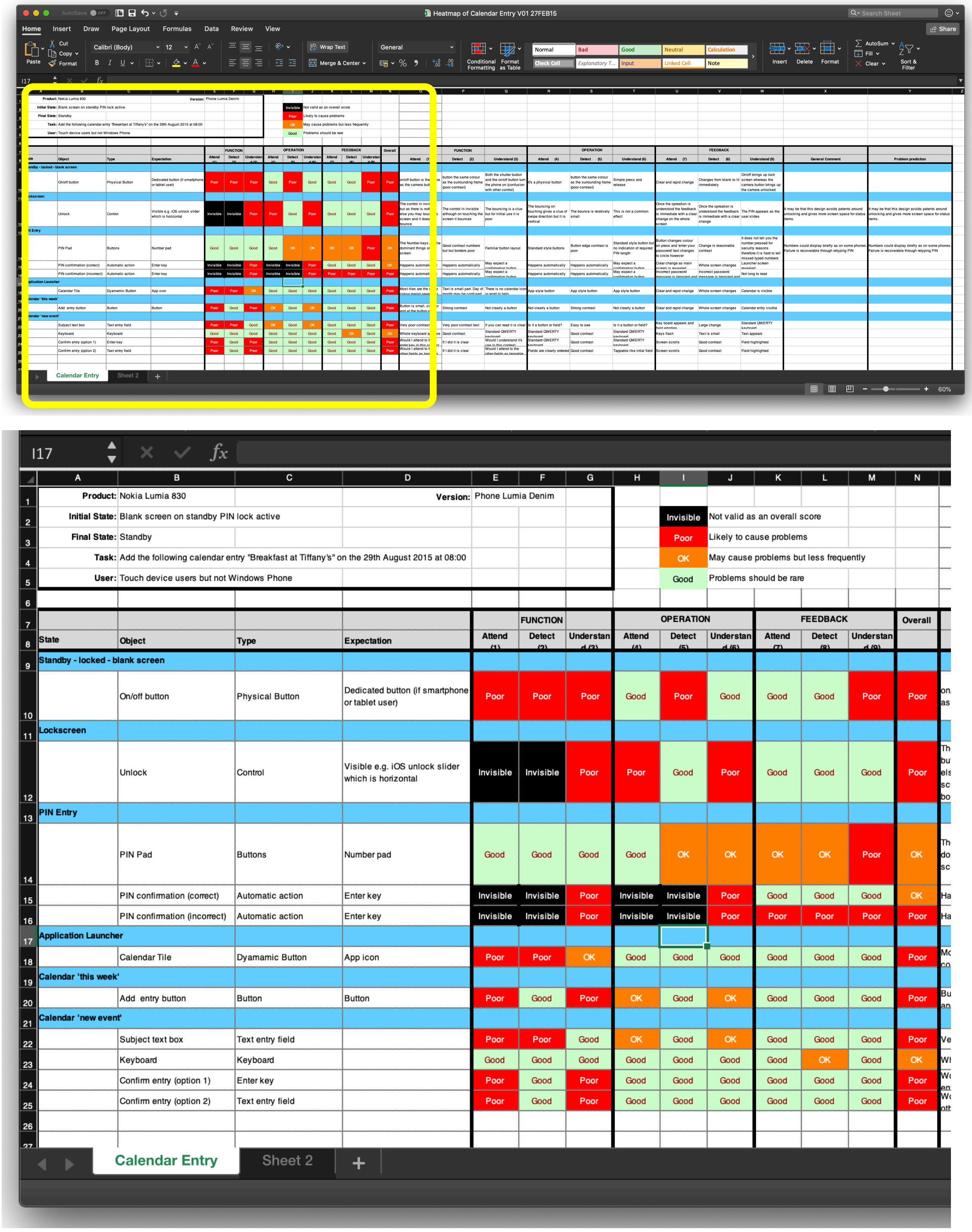
In this section three key elements for the UEM have been described, namely;

1. The underlying detailed model based on a staged 'hurdle' metaphor (Figure 5-4)
2. The hurdle questions to help evaluate each hurdle (Table 5-5)
3. A rating scale (Figure 5-7)

This leads to the next key challenge of how these elements can best be presented for both recording and feedback to practitioners.

5.7 Presentation

Requirement R1 is to use a common software tool for the recording and presenting of results. The first attempt at this used a spreadsheet which is shown in Figure 5-9.



Product: Nokia Lumia 830 — **Version:** Phone Lumia Denim
Initial State: Blank screen on standby PIN lock active
Final State: Standby
Task: Add the following calendar entry "Breakfast at Tiffany's" on the 29th August 2015 at 08:00
User: Touch device users but not Windows Phone

Invisible — Not valid as an overall score
Poor — Likely to cause problems
OK — May cause problems but less frequently
Good — Problems should be rare

State	Object	Type	Expectation	FUNCTION Attend (1)	FUNCTION Detect (2)	FUNCTION Understand (3)	OPERATION Attend (4)	OPERATION Detect (5)	OPERATION Understand (6)	FEEDBACK Attend (7)	FEEDBACK Detect (8)	FEEDBACK Understand (9)	Overall
Standby - locked - blank screen													
	On/off button	Physical Button	Dedicated button (if smartphone or tablet user)	Poor	Poor	Poor	Good	Poor	Good	Good	Good	Poor	Poor
Lockscreen													
	Unlock	Control	Visible e.g. iOS unlock slider which is horizontal	Invisible	Invisible	Poor	Poor	Good	Poor	Good	Good	Good	Poor
PIN Entry													
	PIN Pad	Buttons	Number pad	Good	Good	Good	Good	OK	OK	OK	OK	Poor	OK
	PIN confirmation (correct)	Automatic action	Enter key	Invisible	Invisible	Poor	Invisible	Invisible	Poor	Good	Good	Good	OK
	PIN confirmation (incorrect)	Automatic action	Enter key	Invisible	Invisible	Poor	Invisible	Invisible	Poor	Poor	Poor	Poor	Poor
Application Launcher													
	Calendar Tile	Dyamamic Button	App icon	Poor	Poor	OK	Good	Good	Good	Good	Good	Good	Poor
Calendar 'this week'													
	Add entry button	Button	Button	Poor	Good	Poor	OK	Good	OK	Good	Good	Good	Poor
Calendar 'new event'													
	Subject text box	Text entry field		Poor	Poor	Good	OK	Good	OK	Good	Good	Good	Poor
	Keyboard	Keyboard		Good	Good	Good	Good	Good	Good	Good	OK	Good	OK
	Confirm entry (option 1)	Enter key		Poor	Good	Poor	Good	Good	Good	Good	Good	Good	Poor
	Confirm entry (option 2)	Text entry field		Poor	Good	Poor	Good	Good	Good	Good	Good	Good	Poor

Calendar Entry | Sheet 2

Figure 5-9: A spreadsheet-based heatmap showing the analysis for creating a calendar entry on Windows Phone 10™

Figure 5-9 consists of the full width of the spreadsheet with an enlarged view of the rating section of the sheet which is highlighted with a yellow rectangle in the complete view. The sheet covers the following elements of analysis:

- The overall context and product (top-left)
- Each task step (Blue rows)
- The interface objects relevant to the task step are listed (rows beneath the task step description)
- For each object, the potential expectation, linked to the prior knowledge of the user, is described. This is deemed to be a precursor to determining comprehension (Adams, Graf and Ernst, 2004).
- The objects are then assessed for 'function', 'operation' and 'feedback' using the traffic light rating scale
- To the right of the ratings is the rationale for each rating

The spreadsheet meets requirement R1 for a cross-platform, widely available tool. It is very flexible in terms of adding and moving rows and can record a significant number of entries. It is relatively weak from a presentation point of view and it is not easy to store screenshots. It could be argued that the spreadsheet gives the impression of rigour and therefore aids 'persuasive power' (R6) but this is offset by its presentational weaknesses. This led to exploring a presentation tool (Microsoft PowerPoint™) as an alternative.

Figures 5-10 to 5-14 are samples taken from the PowerPoint™ version. Figure 5-10 shows this applied to the car HVAC example. There is a compromise, in that the graphical nature means that all the necessary information cannot be presented on a single slide but offers advantages in terms of 'persuasive power'. There are 'before' and 'after' images of the UI. This captures the feedback after the control is operated. It is, therefore, possible to see the visibility issues of the 'function', its 'operation' and the subsequent 'feedback' of operation.

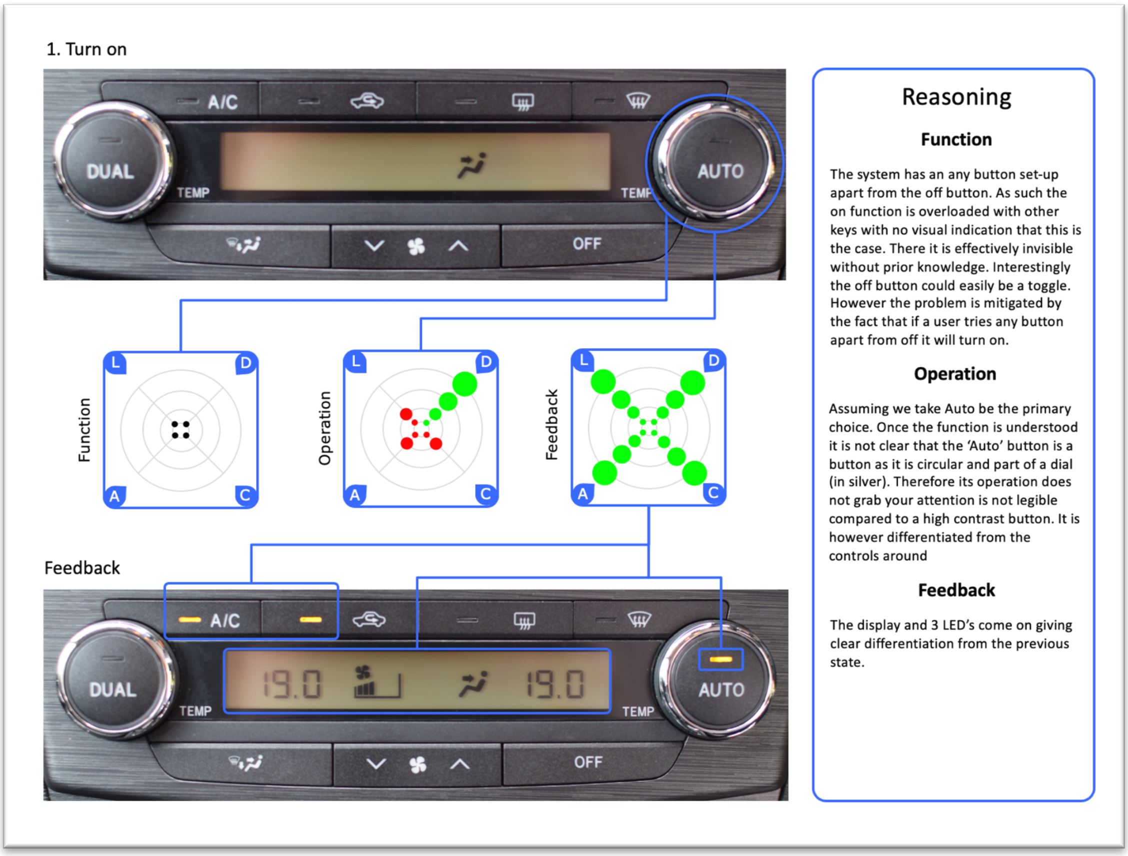


Figure 5-10: Rating and reasoning example for turning on a car HVAC system

In this case, the task step is turning the HVAC on from its off state. There is no visual indication of which button or buttons will turn it on. Therefore, the **'function'** has no overt visual properties hence the black invisible rating shown in Figure 5-10. The **'operation'** of 'turning on' is via the AUTO button (on the right), this scores poorly, as the button is the centre of a rotary dial, this lacks visual clarity concerning how it indicates it is also a button. This leads to poor conspicuity and comprehension. It is clearly differentiated from the other buttons, but overall it scores poorly due to the problems highlighted. However, once **'operated'** the feedback is very clear as 3 LED lights come on as well as various graphics and numbers on the screen. This example shows the power of breaking down the control into **'function'**, **'operation'** and **'feedback'** as they vary from the invisible to highly visible.

In addition to the 'before' and 'after' images of the interface and the ratings, priority is given to showing the reasoning for the rating as ultimately these are subjective, and the rationale is key to understanding why the rating is set as it is. Experience of using the tool shows that it is easy to forget why a rating is chosen that may be the result of some debate at the time. Therefore, recording the rationale proves invaluable when returning to it later.

This format was tried with several different interfaces to see how well it works in terms of layout as well the visibility analysis itself. Examples shown here are for a microwave (Figure 5-11), sports watch (Figure 5-12) and a TV set-top box (Figure 5-13)

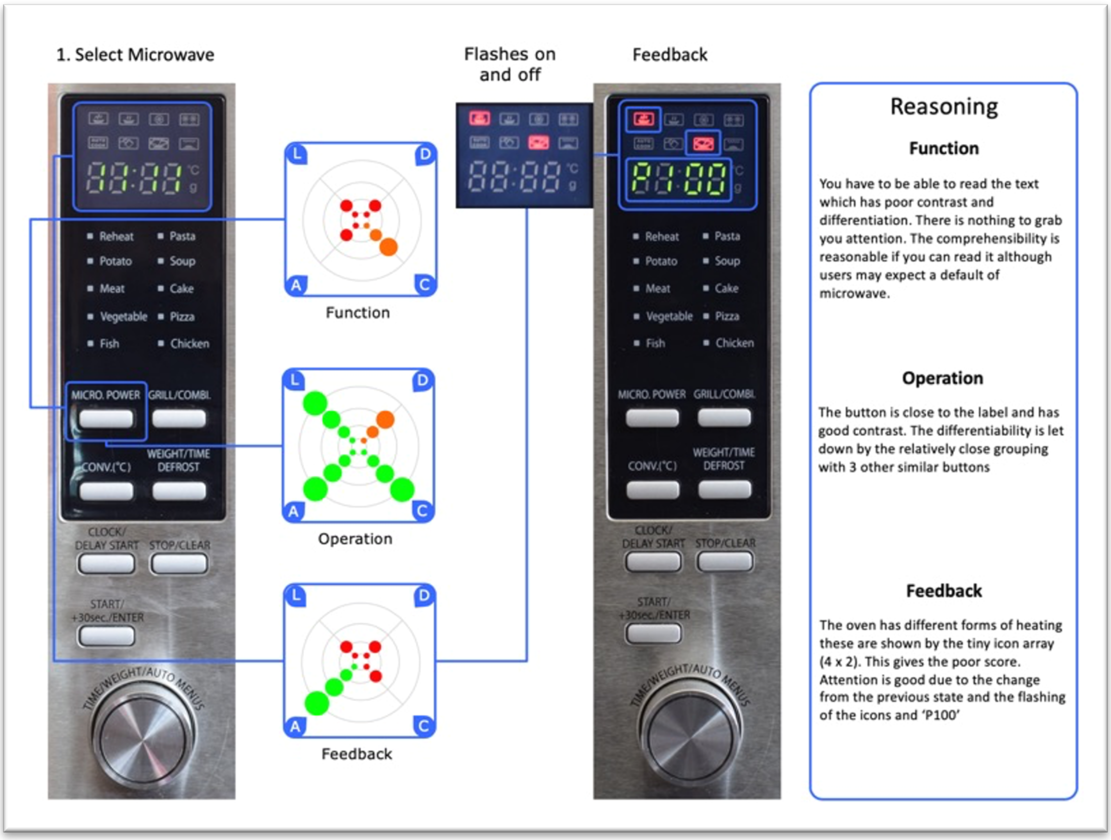


Figure 5-11: An example of rating and reasoning for setting the power level on a microwave

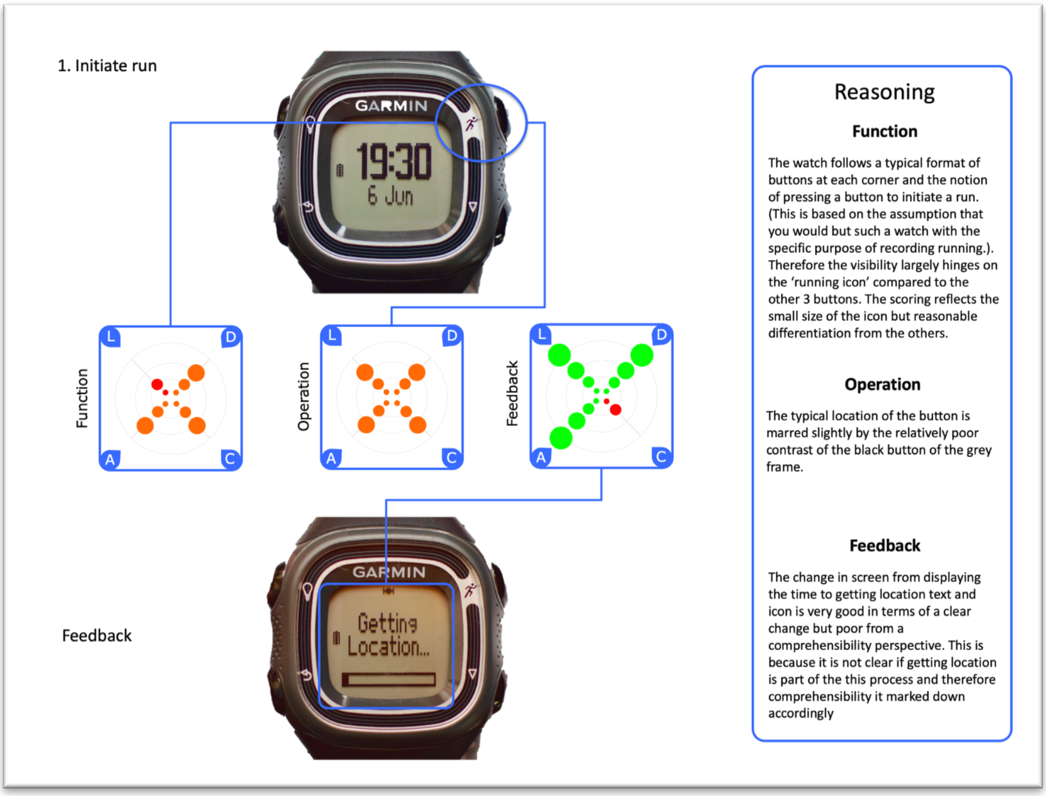


Figure 5-12: An example of rating and reasoning for initiating the recording of a run on sports watch with GPS

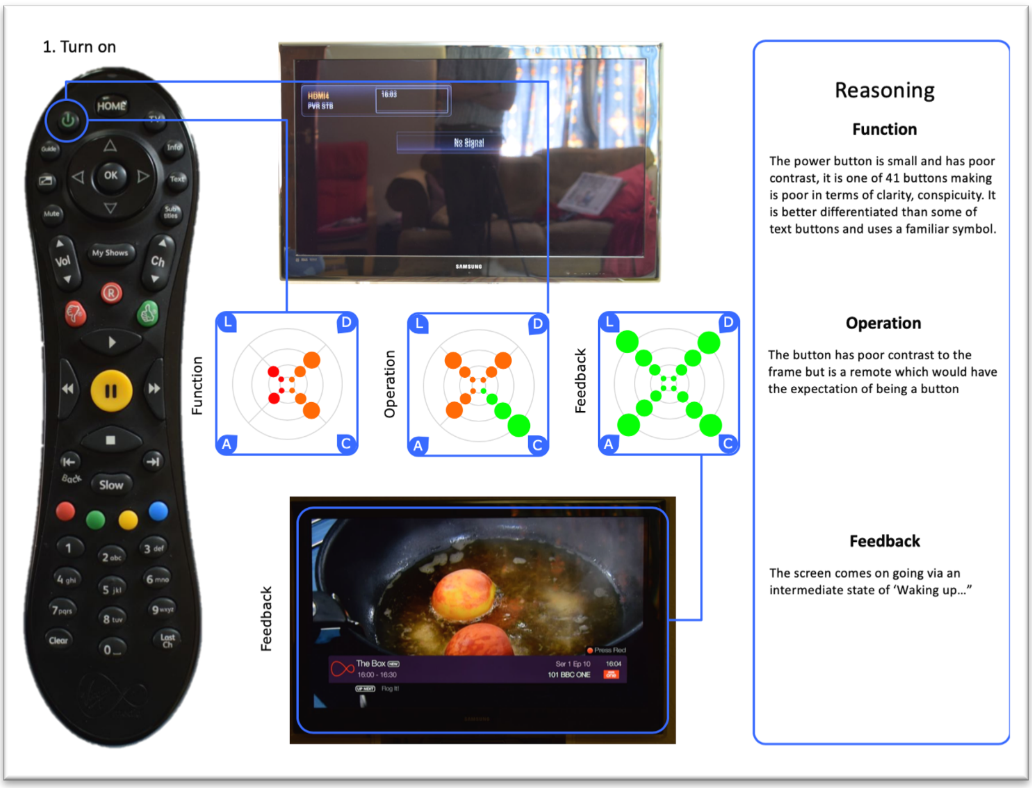


Figure 5-13: An example of turning off a set-top box for a TV

The format copes well in terms of size and format from a watch to a TV. However, the rating display system proved to be relatively slow to update due to the need to change the number and colour of dots in the 'radar' plot. This could be partly overcome by having a selection of different ratings on a 'template' slide that could be copied across, however it was decided to go for the simpler format described in Figure 5-7. The change to this format is shown in figure 5-14 and proved to be much quicker to record the ratings and simpler to read.

Figure 5-14: The Car HVAC with updated rating scales

This rating format includes the full five scales described earlier and elaborated by the hurdle questions in Table 5-5. The example in figure 5-14 shows the value of splitting out 'legibility' into 'focus' and 'clarity' as in this example focusing on the control is easy but the clarity is poor. Conversely, there could be a car dashboard control, for example, that has very good visual attributes, but is let down by being obscured by the steering wheel.

This format was the one used for testing and was applied to the analysis of a mobile phone task that is described in Chapter 6. Although the ratings are a core element of the output, there are several other aspects such as recording answers to the hurdle questions and the predictions of potential usability problems. These are discussed in the next section which covers the end-to-end process for using the tool.

5.8 Tool Process

Having established the core of the UEM tool, based on the detailed model, it is necessary to create an end-to-end process for its use in practice. The outline of the process is described below and shown as a process diagram in Figure 5-15. The numbers in red circles refer to Figures 5-16 and 5-17 that contain thumbnail examples from a full Windows Phone 10™ analysis. This example is the subject of further testing in the next chapter and details about the full analysis and how to access a complete and full-size version are contained in Appendix A.

5.8.1 Scenario Definition

Defining a scenario is key in ensuring the context is understood and the contextual factors can be taken into account. For example, does the scenario cover 'first use' versus 'frequent use' and what type of user does it cover e.g. 'novice' versus a 'power user'. Depending on the complexity of the product, and the number of functions it has will determine if all the product's functionality is analysed or just a subset. Any analysis has to work within the project constraints, such as time and budget. Where it is a subset then it is important to focus on the areas of greatest concern with regards to the likelihood and potential impact of problems and the associated risks. Most products are also likely to have a myriad of potential user journeys and judgement is required in building and selecting scenarios, taking into account the factors outlined above.

The scenario is recorded in terms of:

- Product and version number (Figure 5-16: Image 1)
- User description (Figure 5-16: Image 2)
- Physical environment (Figure 5-16: Image 2)
- Task goal (Figure 5-16: Image 2)
- Analysis notes (Figure 5-16: Image 3)
- Product Overview shot (Figure 5-16: Image 4)
- Eye view showing the product in the wider field of view (This was done using a GoPro HERO4 Silver™ Camera that has a wide-angle of view comparable to human vision [horizontal field of view 122 degrees and vertical 94 degrees (GoPro Inc., 2018) compared to that of human vision of 190 degrees horizontal and 125 degrees vertical (Watson, 2009). An example of this is in Figure 5-16: Image 5.
- A view of the initial state of the product before use (Figure 5-16: Image 6)

5.8.2 Task Analysis

Task analysis (Annett and Duncan, 1967; Shepherd, 1998) is used to break down the task into a series of steps suitable for analysis. The level of granularity (Redish and Wixon, 2002) is key to ensure the step size is at the level of an individual control, where there is a clear 'before' and 'after' from operating the control. An indication that the step size is correct is if it is possible to capture an image of the user interface before and after the control is operated. There are situations where there may be a phased feedback mechanism going through more than one state e.g. via an intermediate progress indicator but the key is the identification and activation of some form of control or control mechanism e.g. a swipe. An example of this is shown in Figure 5-17 image 7.

5.8.3 Record 'before' and 'after' shots

An image of the relevant part or parts of the interface is recorded either via a screenshot or photograph or a combination of the two. All key elements required for the task step need to be recorded so that they are suitable for annotation on a slide. For example, if it is necessary to search for a control within a group of other controls then this context needs to be shown, not just the required control. Figure 5-11 shows just the control panel of the microwave as there is nothing about the door that is likely to impact the panel's use. Figure 5-13 has both the remote control and the TV screen for analysis of a set-top box but not the box itself. In this case, it should ideally include the box for the 'turning on' task as it has red-green LED indicators that assist with this. They would not be relevant to the subsequent tasks and therefore would only be needed for this particular step. This shows the need for both diligence and flexibility when recording what is happening.

5.8.4 Visibility Analysis

Having captured and annotated the 'before' and 'after' images. The 'before' image is analysed with regard to the visibility of the required function/control and its operation. This is done by using the hurdle questions to determine the visibility with regards to the five hurdles. An example of the table used to prompt and record the outcomes is shown in Figure 5-18 and is an enlarged version of image 8 in Figure 5-17. It shows the questions, the answers (rationale) and the rating (score). This is repeated on subsequent slides for operation ('before' image) and feedback ('after' image). Examples of this are shown in Figure 5-17 with images 9 and 10.

5.8.5 Summary Ratings

Having completed the analysis using the hurdle questions. The worst score or rating for the hurdle becomes the overall summary rating for that hurdle. This is on the basis that it represents the highest point of the hurdle (demand) that the user has to overcome. The summary ratings are recorded on the slide with the 'before' and 'after' images. See figure 5-14 for an example.

5.8.6 Usability problem predictions

The process so far, at one level, is about generating the scores and determining the most concerning areas. However, another key part is about making the evaluator think critically and generating insights about potential problems. This leads to the prediction of potential problems. Where a black, red or amber score has been given this is deemed likely to cause a problem and predictions can be made accordingly. An example prediction table is shown in Figure 5-19, which is an enlarged version of image 11 in Figure 5-17. The table includes:

- the predicted problem
- how it relates to one or more of the hurdles
- the predicted probability (very low, low, medium, high)
- the predicted level of impact (high, medium and low)
- the overall rationale for the prediction

This analysis process is repeated for each task step and 'global functions' are considered, which are discussed in the next section.

5.8.7 Global Functions

Many user interfaces have functions that are available all or most of the time, regardless of the state of the UI, and are given the description 'global functions'. These include things such as an on/off button or in the case of a Windows Phone 10 [running on a Nokia 830™] three permanent keys at the bottom of the phone (back, home and search) which are shown in image 12 in Figure 5-17. This is important because they may be used incorrectly by the user and represent potential sources of usability problems.

On more complex interfaces a judgement has to be made of what to include in the analysis if there are many such functions. For example, with a car do you include the brake, clutch and accelerator pedals which are always available, for tasks not directly related to the control of the vehicle? The simple answer may be no but more modern cars such as the Toyota Avensis™ product example requires operating the clutch pedal and parking brake (Toyota, 2010, p.203) to start the vehicle. Even the steering wheel can impact starting as when it becomes locked the car will not start and requires the steering wheel to be moved at the same time as the start/stop button (Toyota, 2010, p.204).

Care is required not to overlook apparently peripheral elements of the user interface for a particular task step, as they can result in a variety of state transitions beyond the typical or intended user path. For example, with a phone, a voice call could interrupt the flow of some other task that the user may be conducting. Such issues may be included in the analysis of task steps or as part of general global functions.

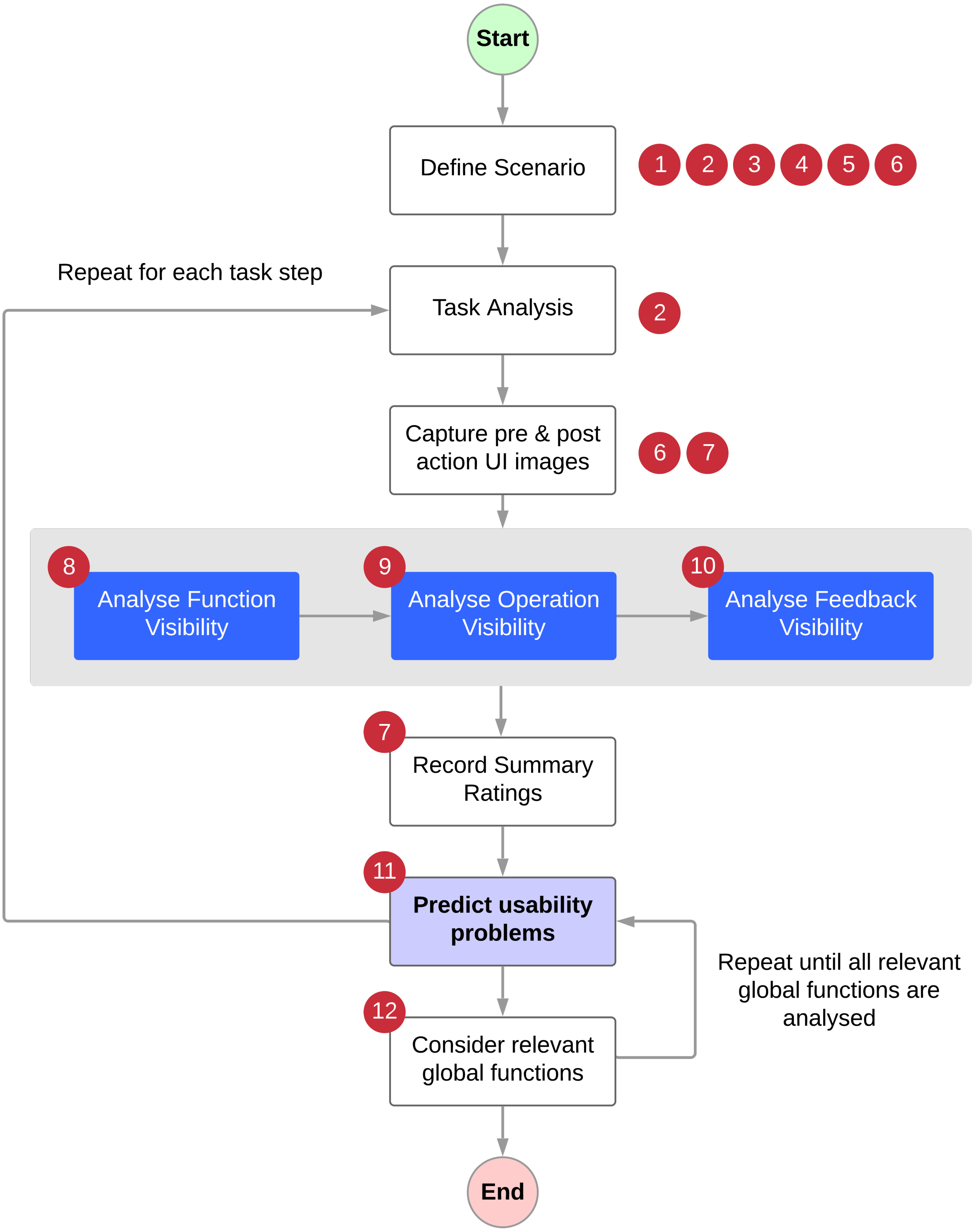


Figure 5-15: End to end visibility audit process [expanded version of Hosking and Clarkson (2018b)]. The numbers in red circles refer to the images in Figures 5-16 & 5-17.

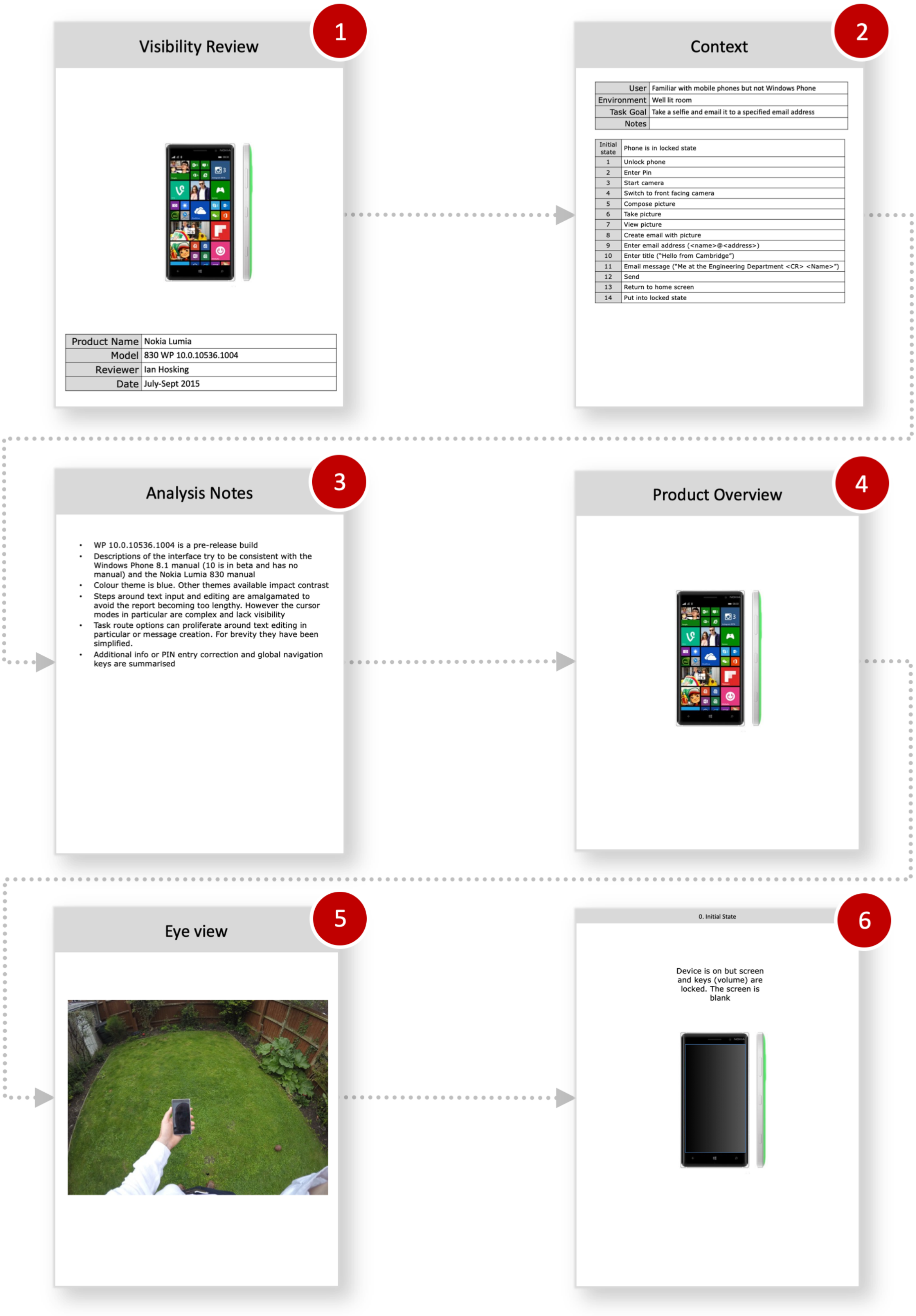


Figure 5-16: Example 'scenario definition' slides from Window Phone 10™ analysis

Figure 5-17: Example 'visibility analysis' slides for first task step Window Phone 10™ Analysis

FUNCTION ANALYSIS: 1. Unlock phone				
Aspect	Question	Answer		Score
Concepts				
	What is the concept of the 'function'?	A physical dedicated lock button (that 'operates' by a short press) N.B. the button is also the power button		
	How is the concept conveyed visually?	Narrow rectangular shape that is proud of the surrounding frame		
	Is it a **familiar** concept to the user? (check against real examples)	Dedicated power+lock buttons are common but not necessarily well understood due to the functional overloading		
	How well is the concept **represented** & are there elements **missing**?	There is no label or symbol to indicate its function. Their is no distinction between 'lock' and 'power'		
	Are there general **variations** of this concept that could cause **confusion**?	Yes. Keys can be overloaded and other phones have it at the top or front. Some keys like this can be sliders.		
Line of sight & focus				
	Does the user have to move to get line of sight?	Yes if looking at the front. Can be done by moving the phone in the hand		
	Can the user focus on it? (bi or vari focal glasses)	Can move the arm to bring it into focus		
Conspicuity				
	Is it in the central visual field?	It is on the side and therefore may not be if the phone is held face on.		
	Is it where the user would expect it to be?	Phone lock buttons have varied locations e.g. the top		
	How many other related elements are there?	1 of 3 on the side (camera & volume)		
	Does it stands out against other elements/background?	The contrast is poor (silver on silver) and there are 2 other similar buttons on the same side		
Clarity				
	What are the key distinguishing features?	**Size**	**Contrast**	
	- Button edge	Okay	Very poor (silver on silver)	
	- Shape	Okay	Very poor (silver on silver)	
Differentiation				
	How different is it from other elements visible at the same time?	There are two other silver on silver buttons with no labels		
	How different is it from other elements visible at other times?	The UI is all on screen with nothing similar		
	Could it be confused with commonly used graphics/symbols that indicate something different?	Plain silver buttons are not associated with specific functions		

Figure 5-18: Example of the hurdle questions table for unlocking the phone

Usability Problem Prediction

Task Step: 1. Turn phone on				
Issue	Visibility problem	Probability	Impact	Rationale
Not finding the power+lock button at all	All	Very Low	High	Very poor contrast but the user would expect there to be a button therefore low probability
Only finding the power+lock button by trial and error e.g. pressing the camera or volume keys N.B. pressing camera key will launch the camera and potentially cause a whole sequence of problems	All	High	Low	It is one of 3 silver buttons that have no labelling
Not pressing the button for long enough leading to rejection that it is the correct button	Concept presentation	Very Low	High	The button press is short (unlike power function) so very unlikely
Confusion between lock & power state i.e. the user may press for too long and go into power off sequence	Concept presentation	Low	High	The power button off state requires a longer press. This may be done accidentally.

Figure 5-19: Example a usability problem prediction table for unlocking the phone

5.9 Conclusion

This chapter has described how a psychophysical understanding of vision can be integrated with a model of human-computer interaction, thus creating a model for UI visibility that addresses the cognitive aspects of vision. Additionally, it also addresses that spectrum of UI visibility issues from the invisible (missed) to things are seen but not correctly understood (misunderstood).

A key aspect of this work was a synthesis of inputs from multiple existing models. This included the cyclical nature of interaction, as seen in the function-operation-feedback breakdown (Neisser, 1976)). Within this cycle, there are a series of stages (Ware, 2003), that are in turn related to a capability-demand model of exclusion (Persad, Langdon and Clarkson, 2007; Waller, Langdon and Clarkson, 2009). This is then conceptualised as a series of hurdles (Nicholl, 2017) which leads to a set of hurdle questions that provide a practical representation suitable for use in evaluation by practitioners.

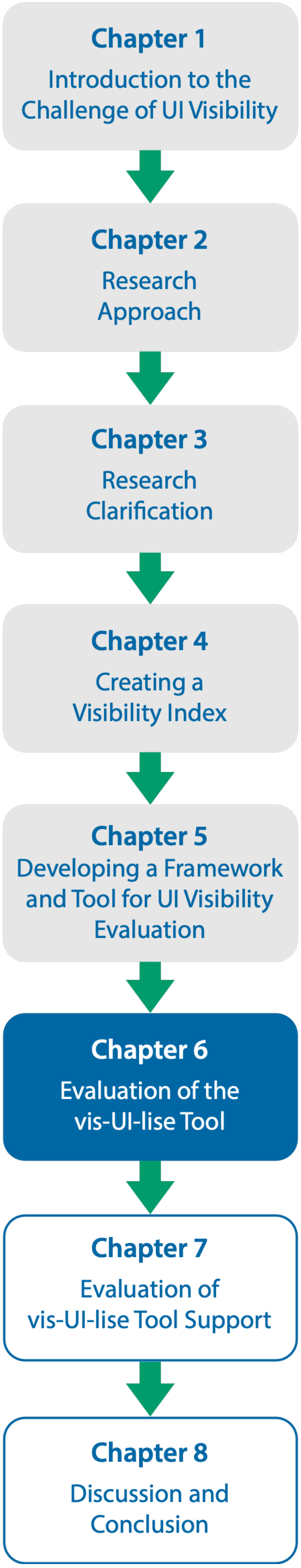
This work makes significant progress in addressing the third initial research question (IRQ) which is repeated below for convenience:

IRQ3: *What can be done to improve it?*

> **SRQ 3a:** *How does a cognitive-based framework for understanding and representing the visibility of user interfaces compare to one based on a simple visibility threshold?*
>
> **SRQ 3b:** *How well can such a framework be embodied in a tool that predicts more usability problems than current approaches?*
>
> **SRQ 3c:** *Can and will practitioners use such a tool to reliably improve the visibility of user interfaces?*

The high-level visibility framework and associated detailed interaction model represent a comprehensive description of UI visibility underpinned with a cognitive-based model of vision. This is the foundation for determining the answer to SRQ 3a, whether such cognitive approach is more effective than one based on a simple threshold. The embodiment of the detailed interaction model in a tool, centred around the hurdle questions, underpins being able to address SRQ 3b. It is also the basis for tackling SRQ 3c with regards to the use by usability practitioners, which is ultimately critical to having a real-world impact regarding improving UI visibility.

With the development of a tool, the key next issue concerns how well the tool performs, which is required to address the 'how well' aspects of IRQ3. This is broken down into two key aspects: its outright performance, which is the subject of the next chapter (6); and how well it works in the hands of usability practitioners which is covered in Chapter 7.


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 6

Evaluation of the vis-UI-lise Tool

6.1 Introduction

The previous chapter described the development of a framework and associated embodiment of a tool for understanding and evaluating UI visibility. This tool was given the name 'vis-UI-lise' as a play upon the word 'visualise' and the abbreviation of user interface, UI. The tool will be referred by this name from now on.

Having established an intervention in the form of the vis-UI-lise tool the next step is the evaluation of the tool. In DRM terms this represents the continuation of the Prescriptive Study (PS) phase with regards to the testing the outright performance of the tool and progression to Descriptive Study II (DSII) for testing it in the hands of usability practitioners. This split in the evaluation and the relationship to the DRM is shown in Figure 6-1 as follows:

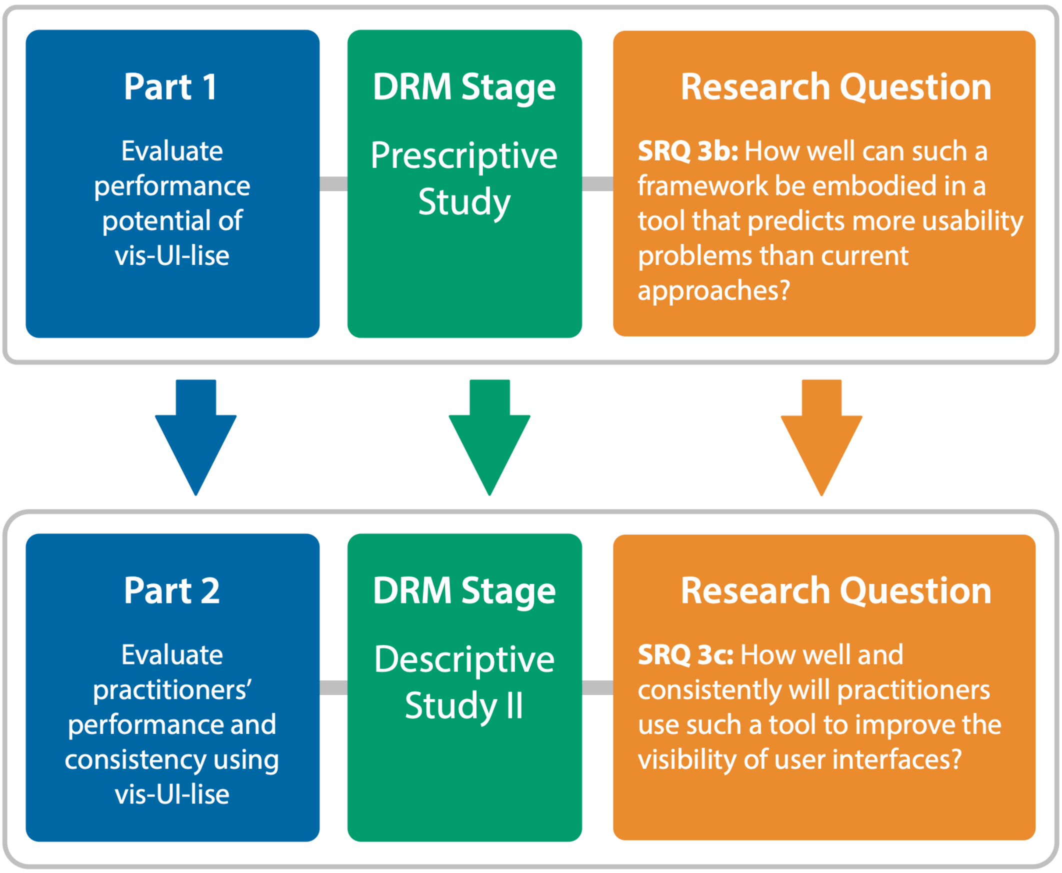


Figure 6-1: Overview of the two-phase approach with the links to the DRM and research questions

6.2 Method overview

This chapter covers Part 1 of the evaluation and the following chapter Part 2. To evaluate vis-UI-lise tool it is necessary to determine what type of tool it is to determine an appropriate evaluation approach. It fits within the broad category of usability evaluation methods (UEMs) and can be sub-categorised as an inspection tool combining elements of a 'cognitive walkthrough' and an 'expert review' (see Table 3-10 in Section 3.9 of Chapter 3 for a list of UEM types). The focus of the tool is the identification of potential usability problems in advance of user testing. Therefore, the approach to the evaluation of the tool has to be appropriate to the type of tool that it is.

The evaluation is performed by the researcher and developer of the tool. It, therefore, represents the performance in 'the hands' of someone who is an expert in the tool. It should represent the upper performance achievable and therefore a benchmark with which comparisons can be made.

6.2.1 Methodological Background

The evaluation of usability evaluation methods has similar struggles to those of usability and usability problems discussed earlier in Chapter 3 (Section 3.7 and 3.8). Hartson, Andre and Williges (2003) highlight these problems, namely the lack of standardised definitions, measures or approaches. They provide a comprehensive critique and approach for UEM evaluation. This is complementary and consistent with the work of Blandford et al. (Blandford et al., 2008) used earlier to outline the requirements for a UEM. The research described above also addresses the concerns raised by Gray & Salzman (1998) about the lack of robust experimental method.

One approach is to evaluate against a 'standard usability problem set', in other words, a list of all the actual usability problems that exist for a particular product. This represents a reference 'benchmark' that allows a comparison of the outputs across different UEM tools to determine their relative performance. This approach is outlined in Figure 6-2 showing how two different UEMs (A & B) can be compared. The 'standard usability problem set' sits at the centre of this diagram.

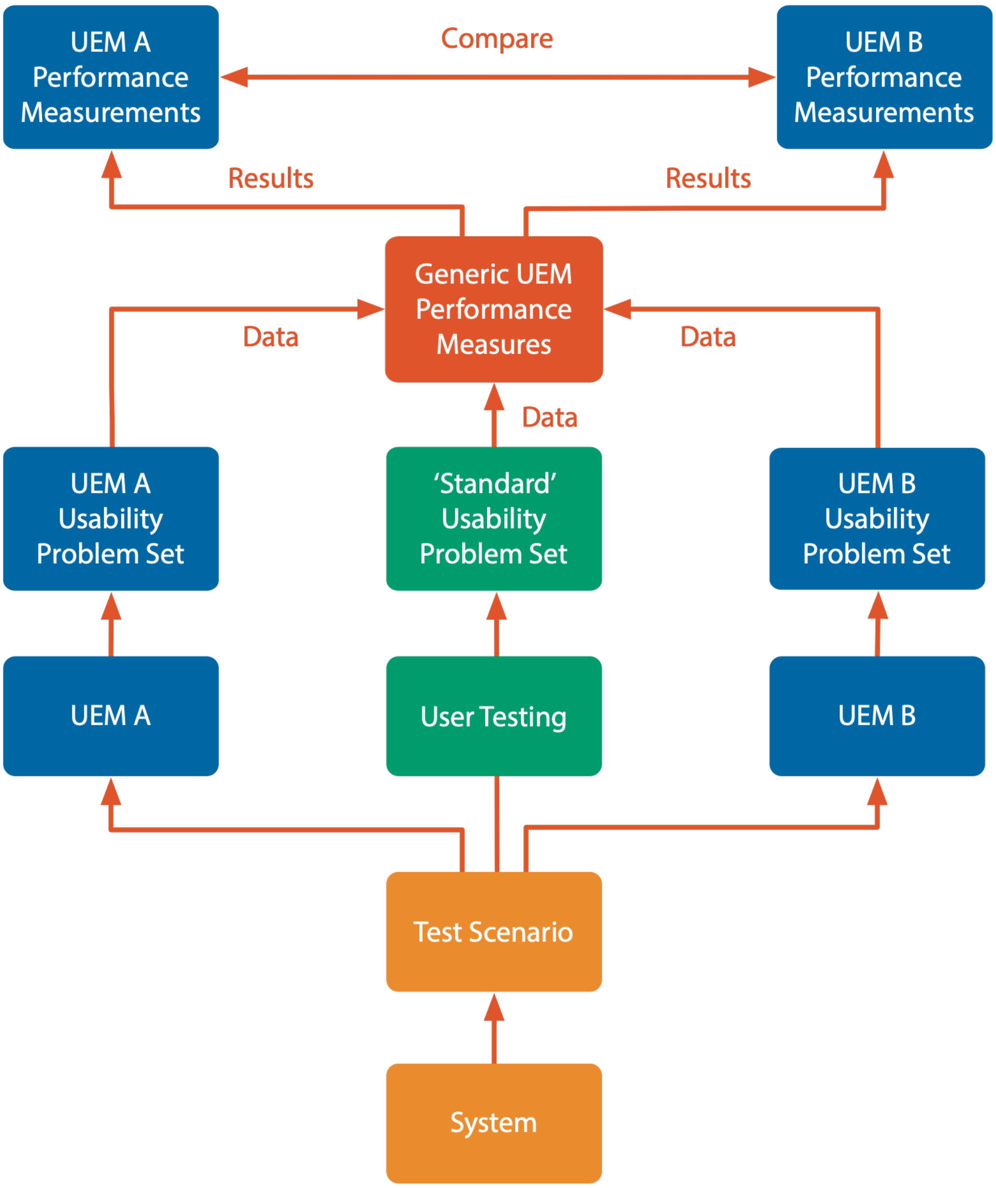


Figure 6-2: A generic approach to comparing the performance of different UEMs by reference to a standard set of usability problems

Determining the standard set is problematic. Figure 6-2 shows this being generated via user testing, based on a defined test scenario (see the items in orange). This represents one option for creating a set and Hartson, Andre and Williges (2003) describe four potential approaches, where items 3 and 4 are two different forms of user testing:

1. Seeding with known usability problems
2. Union of usability problem sets from the UEMs being compared
3. Laboratory-based testing
4. Asymptotic Laboratory-based testing

Deliberately introducing usability problems, 'seeding', is unlikely to replicate the nuances of real-world systems and is predicated on doing this to a system that does not have any usability problems, which seems improbable. The second option is appealing as it requires little additional work, assuming the product has already been tested with other UEMs, as it is the combination of such tests. However, it is weak from a validity perspective as it is not independent of the data. The third, laboratory testing, is deemed to be the 'gold standard' (Landauer, 1995, p.281) but suffers from the 'synthetic' environment of the laboratory where is it hard to replicate real-world conditions. Regardless of these issues, laboratory testing also runs the risk of missing usability problems. This is in part due to the variability across users and leads to the question of how many users do you need to test with to overcome this variability? Addressing this issue also leads to the reason for option 4 in the list, which we will return to later.

This 'how many' question has taxed usability researchers over the years. It has been found that an accumulative binomial probability formula provides a good fit with the problem discovery rates in usability testing (Virzi, 1992; Nielsen and Landauer, 1993; Turner, Lewis and Nielsen, 2006). The formula is as follows:

$$Proportion\ of\ unique\ problems\ found = 1 - (1 - p)^n$$

Where 'p' is the problem discovery rate and 'n' the number of users (test subjects). Determining the 'problem rate' is problematic with different values being reported, for example, Hartson et al. (2003) report values ranging from 0.16 to 0.42. The impact of the 'p' value in the number of problems that will be discovered is shown in Figure 6-3 with 'p' values of 0.01, 0.1, 0.15, and 0.3 plotted.

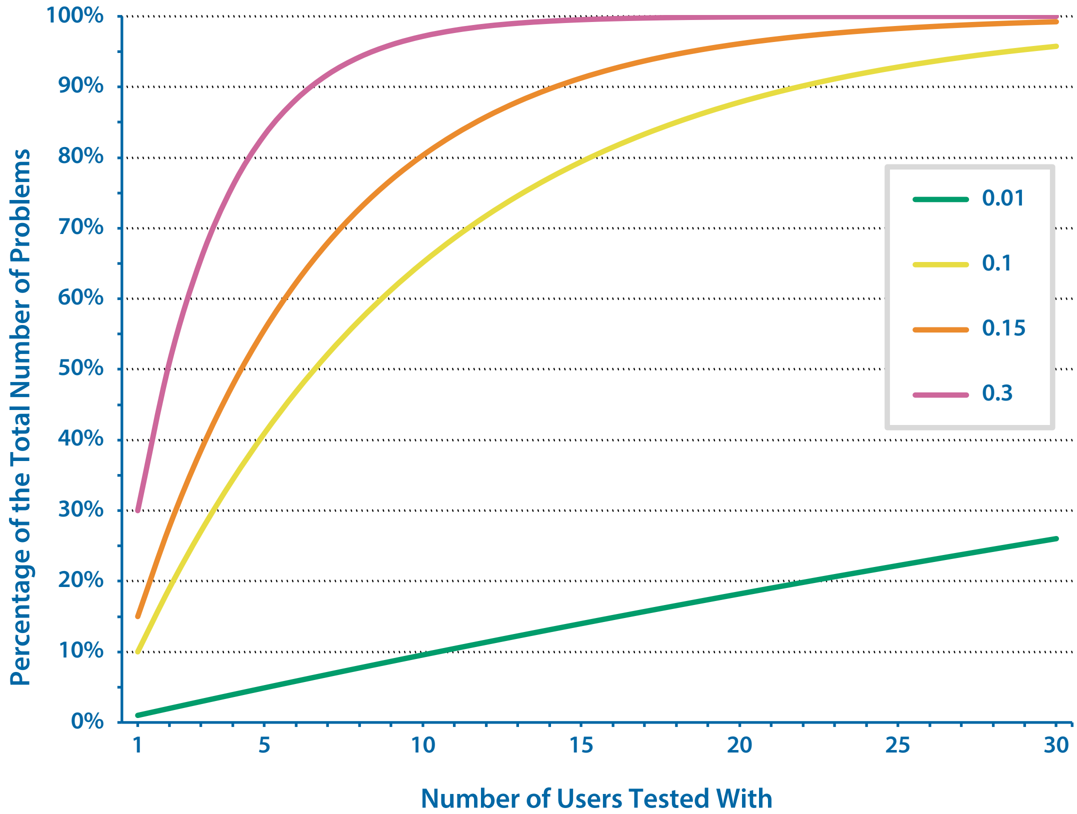


Figure 6-3: Problem discovery curves at different rates of discovery

The oft-quoted "five users find 85% of problems" (Borsci et al., 2013) is based on a 'p' value of 0.31 (Nielsen and Landauer, 1993) which itself is the mean of 13 values from the study of 11 different systems with 'p' values varying from 0.12 to 0.58.

There has been considerable debate about the problems with the formula and various adaptions proposed (Faulkner, 2003; Turner, Lewis and Nielsen, 2006; Borsci et al., 2013). One of the key issues is the assumption that all problems have the same probability of occurring, yet the reality is that problem probability will be heterogeneous and corrections have been proposed to address this (Schmettow, 2008). However, for this work, the basic mathematical model is taken as a starting point for understanding the likely nature of the discovery rate and therefore the potential size of the problem set. This is in-line with Hartson, Andre and Williges (2003) asymptotic laboratory-based testing approach (the 4th item in the list of approaches described earlier), in other words, it is possible to see when the 'full set' size is being reached i.e. it levels off asymptotically.

Having established an approach to characterising a reference problem set this then opens up a range of measures for the performance of a UEM. This is shown in red towards the top-middle of Figure 5-2 and the individual measures are described in the following section.

6.2.2 UEM Performance Measures

The first two measures build on the earlier work of Bastien and Scaping (1995) and Sears (1997) and are as follows:

Thoroughness

Thoroughness is defined as the proportion of real problems found using the UEM to the number of problems known to exist in the product being evaluated, as provided by the standard usability problem set. It produces a value between 0 and 1, where 1 represents perfect thoroughness i.e. all the real problems, found or predicted, match the number of real problems that exist in total.

$$Thoroughness = \frac{Number\ of\ real\ problems\ found}{Number\ of\ real\ problems\ that\ exist}$$

Validity

Validity is the proportion of problems, predicted or found, by the candidate UEM that are real usability problems.

$$Validity = \frac{Number\ of\ real\ problems\ found}{Number\ of\ issues\ identified\ as\ problems}$$

Again, this produces a value between 0 and 1. Where 1 indicates that all the problems ‘identified’ are ‘real’.

Effectiveness

It becomes apparent that it is possible to have a high level of ‘thoroughness’, but the predicted problem list may also to contain a high number of problems that are not real i.e. low ‘validity’ and vice versa. To address this Hartson et al. (2003) introduced the notion of ‘effectiveness’ which is a figure of merit (0 to 1) based on the product of ‘thoroughness’ and ‘validity’.

$$Effectiveness = Thoroughness \times Validity$$

Reliability

Reliability refers to the consistency of usability test results across different evaluators using the UEM. As previously mentioned, there is a significant problem with inter-rater reliability (See Chapter 3, Section 3.8) and therefore this is an important measure. Hartson et al. (2003) outline several different statistical approaches for calculating this, such as the Pearson correlation coefficient (r), Cohen's kappa coefficient (K) and Kendall’s coefficient of concordance (W).

Downstream Utility

Ultimately the value of a UEM comes down to its ability to enable positive change to a user interface to improve its usability. The requirements derived earlier in Chapter 5 from the work of Blandford and Green (2008) and Blandford et. al (2008) help inform what needs to be aimed for in this regard. In particular the notion of 'persuasive power' with the project team and the integration of outputs into the design process to make it part of design practice. As such, this is a more complex, multifactorial, phenomenon to measure and lends itself more readily to a qualitative evaluation. The issue of downstream utility was inherent in the subsequent work with practitioners, which took a qualitative approach to understand it. This work is covered in Chapter 7.

Cost-effectiveness

Cost-effectiveness can be a direct financial comparison between different UEMs or against testing with real users. Cost-benefits can be split to those associated with the development process and those that occur 'downstream' either as a result of changes made or problems missed (Mantei and Teorey, 1988; Karat, 1990; Nielsen, 1993b).

Savings during the development process focus on the cost of change and how this can escalate dramatically the later the change is made. This is because changes are typically easier and quicker to make in the early stages e.g. simple prototypes can be changed rapidly, as opposed to fully developed prototypes where significant amounts of software code may need changing. Boehm and Papaccio (1988) cite examples where the cost of change increases by a factor of 50 to 200 in the later stages versus the earlier ones. Although software development practices have changed significantly since then, with a greater emphasis on iteration (Boehm, 2006), it remains a significant concern and one where early usability evaluation can spot problems and reduce the cost of change. The vis-UI-lise tool is well suited to evaluating user interfaces at an early stage of development and therefore can help reduce the risk of costly changes in the later development stages.

Once problems have been identified and fixed, they produce downstream operational savings. Examples of this include:

- Reduced training costs
- Reduced support costs
- Reduced product returns
- Increased productivity

Deane et al. (2011) have produced a spreadsheet tool for systematically developing a comprehensive business case for the application of usability. This covers a broad range of cost-benefit issues across the development and operational phases. As with 'downstream utility', this aspect of UEM performance is covered in the work with practitioners covered in Chapter 7.

6.3 Describing Usability Problems

Having defined a set of UEM performance measures it is imperative that there is a reliable approach to ensuring the consistent, accurate and complete description of usability problems. From Chapter 3, Section 3.8 three core guiding works were used to form such an approach.

The first is from the work of Lavery, Cockton, and Atkinson (1997) which includes a helpful model of a usability problem which is broken down into 'cause', 'breakdown' and 'outcome'. This is shown in Figure 6-4 below.

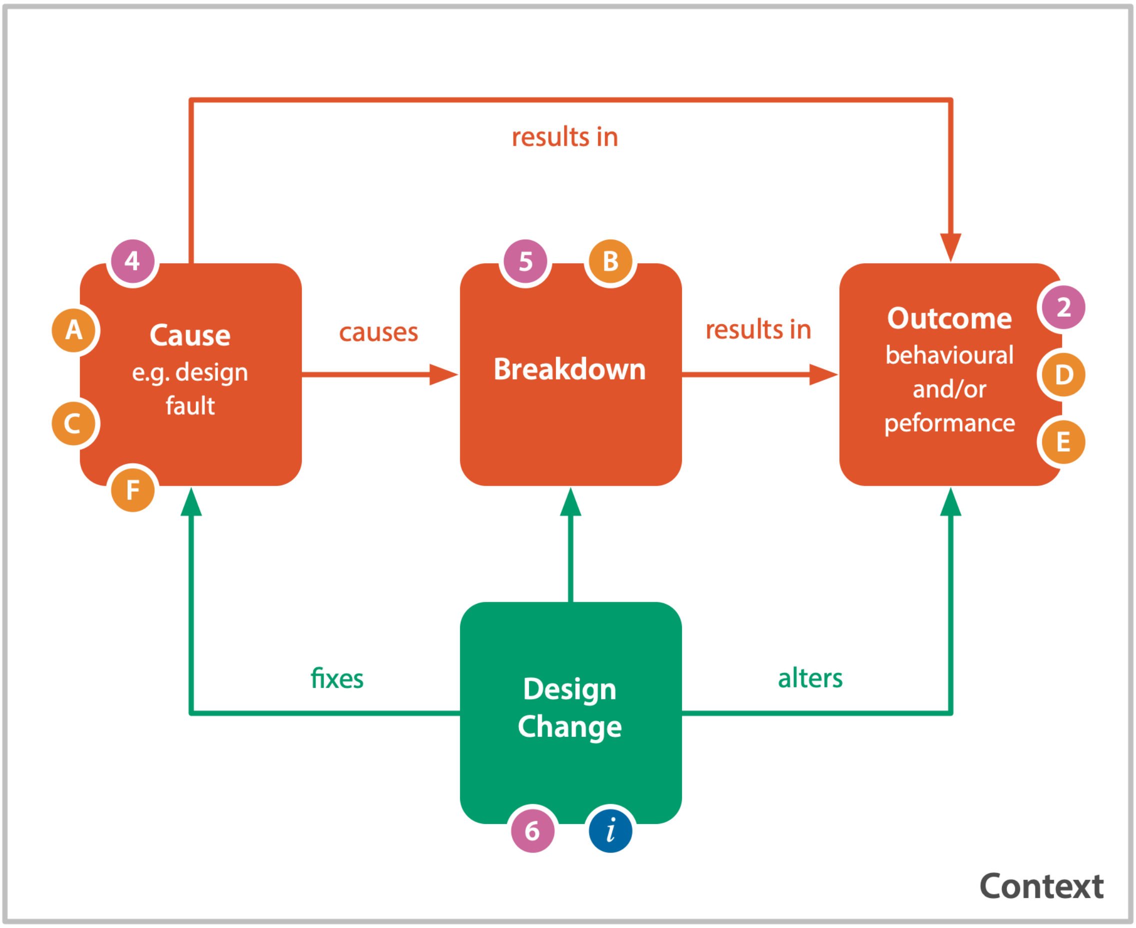


Figure 6-4: Generic model of a usability problem based on Lavery, Cockton, and Atkinson (1997) with the additional numbers to show mapping to Capra and Smith-Jackson's (2005) 10 guidelines (see Page188), letters show problem description attributes used in the vis-UI-lise tool (see Page 190) and a Roman numeral for the additional attribute for user testing

The paper includes examples of these three entities which are summarised in Table 6-1 as follows.

Table 6-1: Examples of 'breakdowns', 'outcomes', and a 'cause' from Lavery, Cockton, and Atkinson (1997)

Example breakdown types
the user forming an inappropriate goal
the user selecting an inappropriate action
the user not perceiving the feedback
the user misinterpreting the feedback
Example outcome types
the user's task failed
the user's performance time was increased
the user's quality of work suffered
Example Cause
For example, a design fault such as a button lacking salience may result in the user not finding the correct action to achieve their goal.

The second work is that of Capra and Smith-Jackson (2005) that contains 10 guidelines as follows:

1. Be clear and precise while avoiding wordiness and jargon
2. Describe the impact and severity of the problem
3. Justify the problem with data from the study
4. Describe the cause of the problem
5. Describe observed user actions
6. Describe a solution to the problem
7. Consider politics and diplomacy when writing your description
8. Describe your methodology and background
9. Help the reader sympathize with the user
10. Be professional and scientific in your description

Guidelines 2, 4, 5, and 6 are mapped to Lavery, Cockton, and Atkinson's (1997) model in Figure 6-4 to show the consistency across the two approaches.

Thirdly and finally, Keenan, Hartson and Kafura (1999) is used to provide a detailed taxonomy of problems that broadly map to the 'causes' from Lavery, Cockton, and Atkinson's (1997) model. The breakdown of these is shown in Figure 6-5.

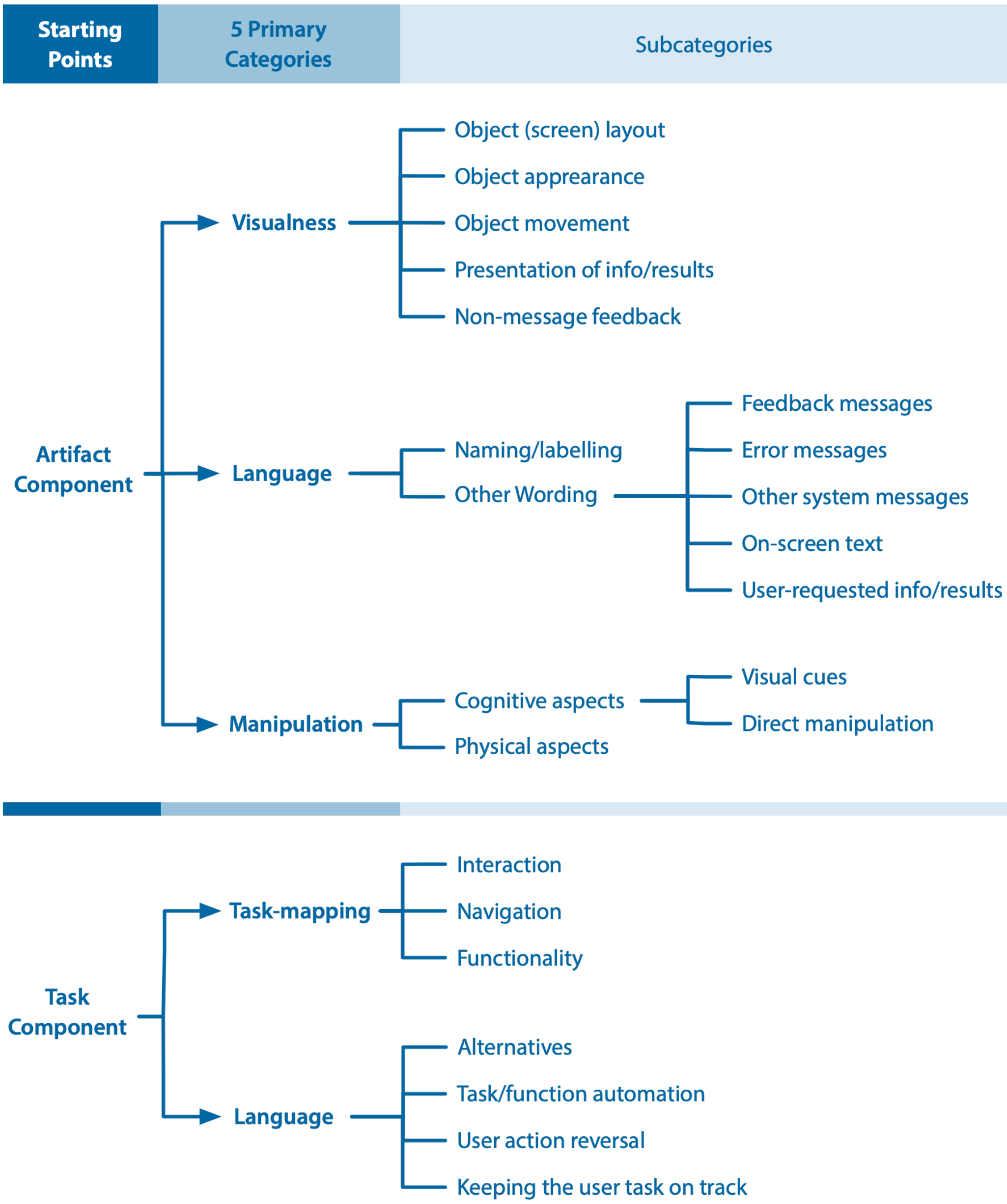


Figure 6-5: Usability Problem Taxonomy categories redrawn from Keenan, Hartson and Kafura (1999)

This led to the following problem description breakdown targeted at (1) the description of predicted problems within the vis-UI-lise tool; and (2) for the analysis of problems observed during user testing that produced the 'standard usability problem set' to evaluate the vis-UI-lise tool against.

For use within the vis-UI-lise tool

For each task step, the interface is analysed in terms of function, operation and feedback using a set of hurdle questions (see Table 5-5). This represents, to a large degree, the 'cause' within Lavery, Cockton, and Atkinson's (1997) model and is marked as 'a' on in Figure 6-4. Based on the output of this a number of predicted problems were described using the following attributes (see Figure 5-19 for an example) which are labelled 'b' to 'f', following on from the 'a' described above, and shown in Figure 6-4.

b) Issue description
c) Visibility problem (a simple 'yes' or 'no')
d) Probability of it occurring
e) Impact if does occur
f) Rationale for why it might be a problem

For use with the analysis of user testing

For user testing, the description was modified to reflect the fact that it is based on real instances of problems as opposed to predictions. This includes the removal of 'd', probability. The rationale for the prediction, 'f', was replaced with 'fix as check', or in other words a proposed design change that represents a check on whether the problem was described in way that could lead to an appropriate solution. This is shown in Figure 6-4 as 'i'. This is in line with the principles behind describing problems proposed by Capra and Smith-Jackson (2005).

We will return later to the reality of 'problem description' in section 6.7.2.

6.4 Protocol for vis-UI-lise tool testing

Having established an overall approach for evaluating the vis-UI-lise tool it is necessary to translate this into a specific experimental protocol. The first thing to address is the problem of performing a direct A-B comparison as outlined in Figure 6-2, i.e. comparing vis-UI-lise to a similar UEM. Testing one UEM after another leads to a very strong order effect because whichever tool is used first, will uncover usability problems that will heavily influence using the second UEM, as they will be in the mind of the evaluator. In other words, they will have seen many of the problems before and be sensitised to them as they use the second UEM. A way around this is to use a different evaluator for the comparator UEM, however, this opens up the problem of inter-rater reliability (Hertzum and Jacobsen, 2003b; Hertzum, Molich and Jacobsen, 2014). This could be addressed by recruiting multiple evaluators, but this has a significant time/cost penalty that is prohibitive for an early-stage evaluation of a new tool.

Therefore, for the first iteration of testing an alternative approach is to compare the vis-UI-lise tool against other documented UEM performance measurements already in the literature. This is a fair compromise as the out-and-out performance of the vi-UI-lise predictions versus actual problems observed is still measured, as well as relating it to other UEM performance results. The modified procedure is shown in Figure 6-6 which can be compared to the direct A-B comparison one shown in Figure 6-2. This approach is performed in two stages. The first stage is the use of the vis-UI-lise tool to produce a 'vis-UI-lise Predicted Problem Set' shown in blue. The second stage is to test with real users to produce an 'Observed Usability Problem Set' shown in green. The data from these are then used to calculate the 'vis-UI-lise Performance Measures'. These measures can be used to compare with other UEM measures available in the literature, which is shown in pink.

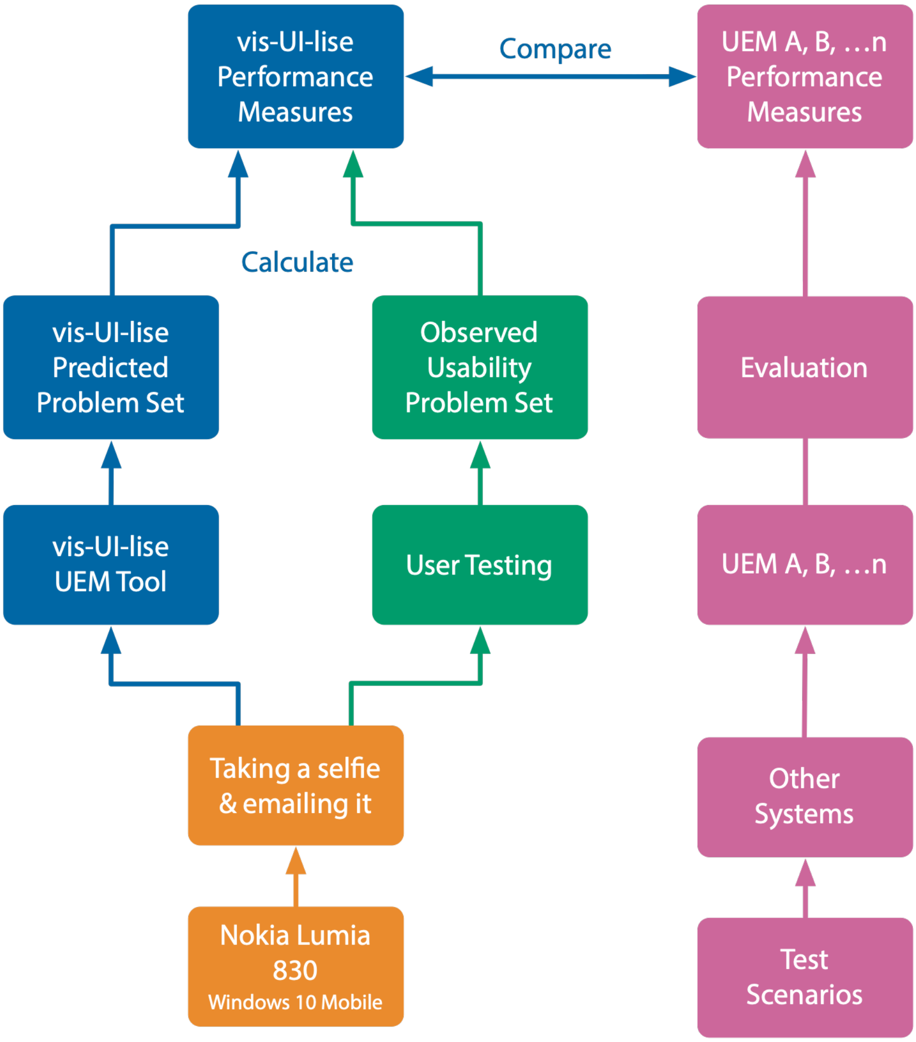


Figure 6-6: Specific approach to evaluating the vis-UI-lise Tool with regards to outright performance based on Figure 6-2

A Nokia Lumia 830™ running Microsoft Windows 10™ Mobile was chosen on the basis that it would be easier to find participants who had not used this particular mobile operating system. This was because it had a very small market share compared to the dominant offerings from Apple's iOS™ and Google's Android™. In terms of the task, this was to take a 'selfie' and email it to a specific email address. This was chosen because they represent common tasks that participants are likely to be conceptually familiar with, but also stretch the range of interface elements that are required e.g. switching from the rear to front-facing camera and having to attach the photograph as part of an email. This is shown at the bottom-left of Figure 6-6 in orange.

This test scenario was broken down into a step-by-step task sequence which is shown in Table 6-2 as follows.

Table 6-2 – Step by step task breakdown for the chosen task goal

Initial state: Phone is locked	
1	Unlock phone
2	Enter Pin
3	Start camera
4	Switch to front-facing camera
5	Compose picture
6	Take picture
7	View picture
8	Create email with picture
9	Enter email address (<name>@<address>)
10	Enter title ("Hello from Cambridge")
11	Email message ("Me at the Engineering Department <CR> <Name>")
12	Send
13	Return to home screen
14	Put into locked state

Each of the task steps was analysed according to the vis-UI-lise tool process that is described in Chapter 5. A summary of the process from this chapter is shown in Figure 6-7. For each of the task steps the 'before' (used for the analysis of function and operation) and 'after' screen image (used for the analysis of feedback) was captured to aid the evaluation. It should be pointed out that in the function-operation-feedback cycle, typically the 'after' or feedback image becomes the 'before' image for the next task step.

This was recorded in Microsoft PowerPoint™ and the saved as a PDF version to avoid accidental editing, in other words, a 'fixed' final output for comparison with the user testing. This is shown in the upper left-hand side of Figure 6-6.

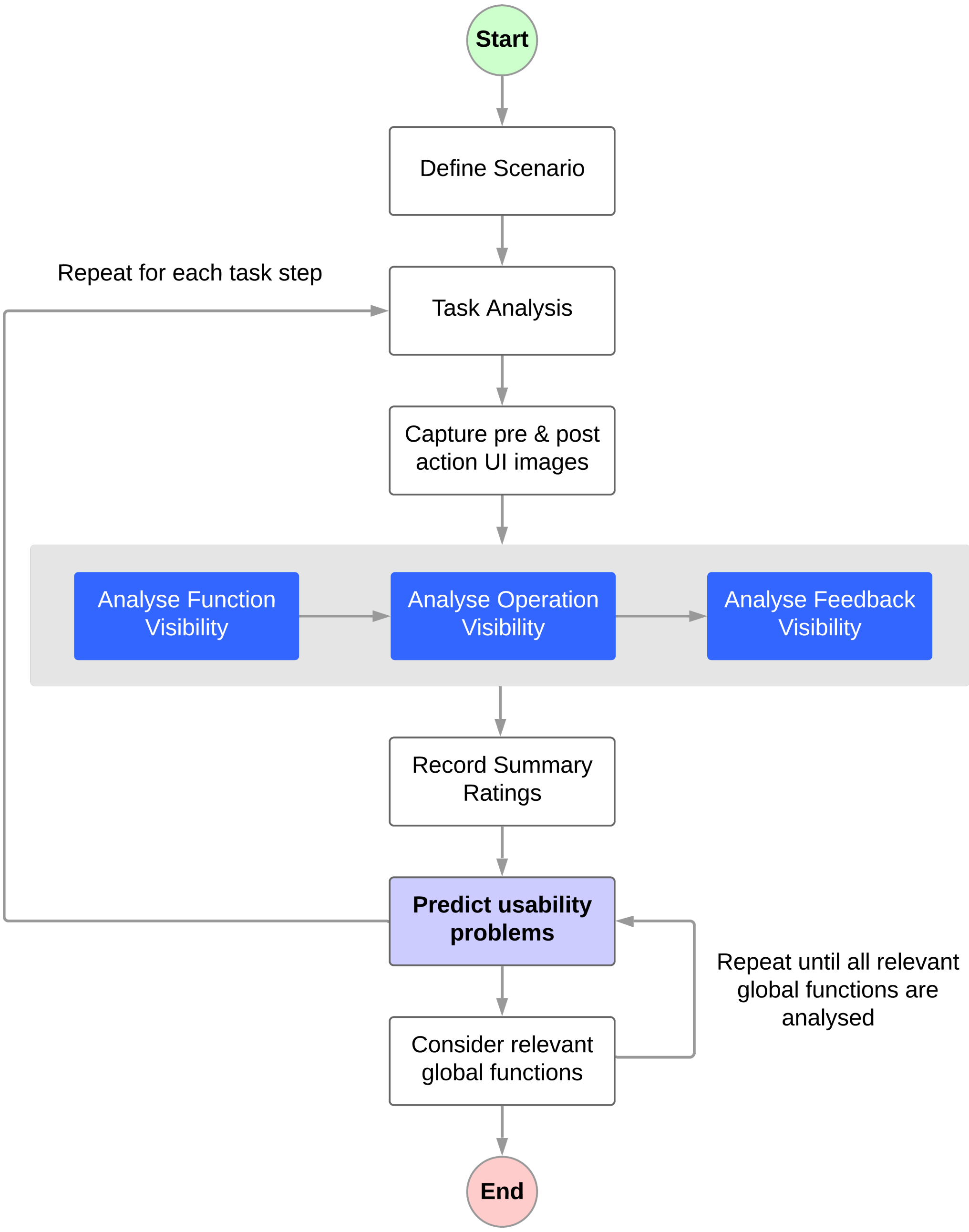


Figure 6-7: A summary of the vis-UI-lise process from Section 5.8

6.5 Protocol for User Testing

The scenario used for testing the vis-UI-lise tool becomes the basis for the user testing to allow comparison between predicted and actual results.

6.5.1 User testing participant selection

The selection of participants (sampling) is something that is not majored on in the debate about problem discovery rates (see section 6.2.1), which seems an odd omission considering its importance. A typical starting point for selection, in a study, is considering a statistically representative sample of a target population. However, here the focus is not how people perform across a defined population but selecting people who will help find as many problems as possible. An approach for 'finding' such people is to look for 'edge cases', sometimes called 'boundary cases' (Keates and Clarkson, 2004; Clarkson, Waller and Cardoso, 2015; Högberg, Brolin and Hanson, 2015). These are people at the edges, or boundary, of the target user base for some reason. This can be related to their sensory, physical or cognitive capabilities or other factors such as economic considerations i.e. can they afford the product. Here we interested in their capability. From a usability perspective if you can meet the needs of the boundary cases then everyone within this boundary should be okay. In this instance, we are interested in them because it is reasonable to assume that they are going to struggle more and therefore highlight more problems. Indeed, it may be of interest to go beyond the boundary to participants outside the defined user base as they will come with a different perspective that may more readily highlight usability issues.

The problem of selection is further compounded when it is hard to predict what kind of problems users might encounter before the testing has commenced. Therefore, it might become evident as testing progresses that edge cases may be different from those originally envisioned e.g. relating to particular prior experiences that may help or hinder the participant.

With the problems outlined above a pragmatic approach was to use age, i.e. older people, as a key determinant of being a potential edge case. The background to the impact of age on technology use was discussed in the part of Section 2.2.2. concerning inclusive design. The use of age leads to 'purposeful' sampling, which is succinctly described by Palinkas (2015) as *"a technique widely used in qualitative research for the identification and selection of information-rich cases for the most effective use of limited resources"*. The 'identification' was aided by a simple screener document and was on a 'convenience' basis in that the participants were known directly or indirectly to the researcher. The criteria in the screener were as follows:

- Over 50 years of age
- 50:50 Male to female ratio
- At least some with lower technical proficiency e.g. no or light smartphone use to represent cases right at the boundary or slightly beyond

The target was to recruit 14 participants in total in line with typical problem discovery rates discussed earlier and, in particular, the work of Faulkner (2003). This highlights a meaningful reduction in the standard deviation in the predicted problem set size due to variation of users when there are 10 participants or more.

6.5.2 Ethics

Ethics was guided by the Engineering Department's approach and consistent with the British Psychological Society's (2009) Code of Ethics and Conduct. Of note is the use of an informed consent form that makes it clear that participants have the right to withdraw at any time without having to give a reason. Also, each set of participant's data was anonymised and stored securely by making access via two-factor authentication.

6.5.3 Experimental set-up

A portable recording system was devised to allow rapid and simple set-up in different locations such as the participants' homes. This made it possible to give the participants the option of it being at home or the University of Cambridge. This meant they could choose where they would be most comfortable and to make it feel less like a 'laboratory' experiment. Simple things such as offering a drink at the start were done to help the participant relax.

The set-up consisted of the following items for recording the experiment:

- A GoPro HERO4 Silver™ video camera, which is rugged and small making it suitable for mounting on the body. This was set to a 1440p resolution to allow a framerate of 48 FPS (GoPro Inc., 2015) to ensure rapid actions and transitions with the user interface can be captured. Higher resolutions would result in lower framerates, so this was deemed to be the best compromise between resolution and framerate. The field of view was 94° in the vertical axis and 123° horizontally (GoPro Inc., 2018) allowing a high degree of tolerance to the user moving their arm and hand position.
- The GoPro™ was attached to the participant using a chest strap. Due to the potential sensitivity or discomfort of the location of the camera participants were given the option of not wearing the camera.
- Since the camera was located on the chest it is not possible to easily see the screen on the back or operate the controls. This was overcome by using the remote control and viewing capabilities of the camera. This was done using an Apple iPad™ running the GoPro™ software. This was hidden behind a simple screen (see the left side of figure 6-8).
- There was a second phone that could receive the email from the participant to allow the complete end-to-end operation of the task in situ
- Finally, a structured form was used to write up problems as they were observed. This served the dual purpose of helping with the subsequent

analysis and as a back-up, in case the video set-up failed, or an incident was obscured or not clearly visible from the recording.

The researcher sat adjacent to the participant to allow for a relaxed style during the discussion and to facilitate direct observation while the user was using the phone.

6.5.4 Additional Instruments

In addition to the recording and observation of the participants, a number of other instruments were used to gather data from the study. The aim was to keep the study time to an hour or less to be fair to participants, particularly for the older ones. Therefore, any instruments had to be quick to administer. Three additional ones were chosen as follows:

1. **Retrospective think-aloud protocol** was employed to try to understand more thoroughly the problems that a user faced. In particular, the underlying 'causality' relating to any 'breakdown' that occurred (as per Figure 6-4). Retrospective protocols are surprisingly effective in comparison to concurrent think-aloud protocols. This is thought to be due to the reduction in the load of having to verbalise what the user is experiencing. In studies, a retrospective approach has been shown to produce more detailed answers than concurrent protocols (Ohnemus and Biers, 1993; Van den Haak and De Jong, 2003). Other studies have shown it to have a more modest advantage if any (Peute, de Keizer and Jaspers, 2015), but none the less it is still a very effective technique and appropriate in this case as it does not overload the user or impact their performance while doing the task.

2. To help characterise the participants in relation to them being, or potentially being, boundary cases two well-established questionnaires where used. The first was the **System Usability Scale**, or **SUS** for short (Brooke, 1986). This was used to assess the global usability of the test phone. The SUS questionnaire is a well-validated instrument

(Bangor, Kortum and Miller, 2008) and with only 10 items it is quick to administer. The aim was to understand if the SUS correlated with their actual performance i.e. how many problems they encountered. If it does, then the SUS could potentially be used to screen participants i.e. users that give lower SUS scores for similar products may be more likely to encounter problems.

3. Further characterisation was sought by relating the participant's difficulties to their technology familiarity. Such prior experience is deemed to be a good predictor of usability (Blackler, Popovic and Mahar, 2003; Langdon, Lewis and Clarkson, 2007). Hurtienne, Horn, Langdon and Clarkson (2013) provide a breakdown of the components of prior knowledge and distinguish between 'exposure' and 'competence' measures. Their work shows that competence measures are a better predictor of usability but acknowledge that exposure measures are easier to administer. The **Technology Familiarity Questionnaire** developed by Blackler, Popovic and Mahar (Blackler, Popovic and Mahar, 2003; 2010), an exposure measure, was chosen for this reason. This was modified to include several very specific questions related to operating systems and to confirm that they had not used a Windows Phone. These questions were as follows:

    - What PC operating system do you use most?
    - What Phone operating system do you use most?
    - What Tablet operating system do you use most?
    - Have you ever used a Windows Phone?

    Also, the participants were asked if they had ever taken a 'selfie' before as this is a key aspect of the task. The frequency of taking a selfie was recorded using the scale from The Technology Familiarity Questionnaire i.e. every day to never. Details about accessing a copy of the modified questionnaire can be found in Appendix H.

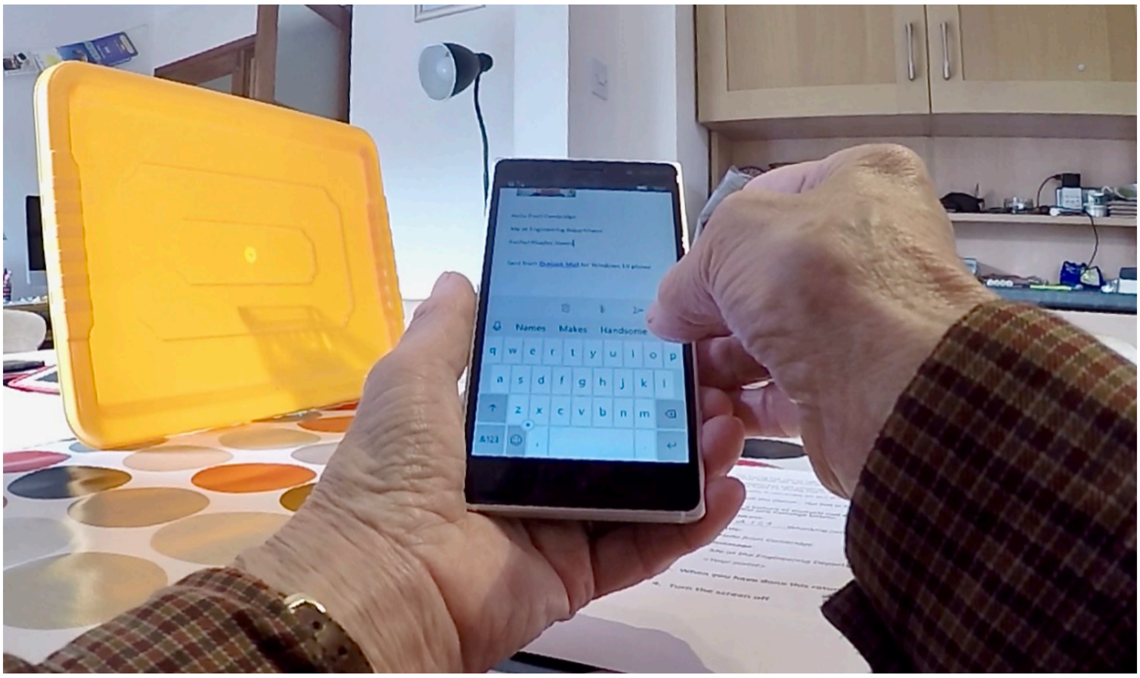


Figure 6-8: A cropped image captured from the video recording using a chest-mounted camera

6.5.5 Protocol Steps

The following protocol, in the form of a printed checklist, was followed to ensure experimental consistency across participants. This formed a series of prompts for all the key parts. The steps were as follows:

1. The age, sex and highest education level was recorded.
2. The purpose of the work was explained, using a simple script to ensure all aspects were covered. The script included the following items:
    - What the research was aimed at i.e. usability
    - What the participant will be doing
    - That it is not a test of them and that there are no right or wrong answers
    - That it will be recorded on video if they are comfortable with it
    - All the results are anonymous

- That they have the right to stop at any time without giving a reason
- That they can ask questions at any time, but that it was important to initially see how they would use it without any help
- They would be prompted to move on if struggling at the point that the issues that were being considered were understood
- There are some questionnaires at the end
- There is an informed consent form to say that they have understood the above

3. The participant reads and signs the informed consent form.
4. The user puts on the chest strap and camera. The recording is started via the remote application and the angle adjusted to ensure effective recording of the phone screen.
5. The participant was given an instruction sheet about the task. This ensures consistency of explanation across different participants. A unique email address is assigned to each participant. This is because the email app will recognise previous email addresses that have been entered and offer it as the first few letters are entered. This ensures that every user gets an identical experience i.e. it is the first time the email address has been entered.
6. The participant was then observed, and notes were taken on a structured form to enable comments to be recorded against the specific task steps shown in Table 6-2. The participant was given the opportunity of looking at the notes at the end to ensure they were comfortable with what had been written.
7. Once the participant had completed the task the researcher asked the participant to talk about specific problems that they had had. The notes taken during the observation stage were used to guide the areas for this retrospective analysis. The key was to focus on things where it was not obvious why they had done it and what the underlying cause might be. Questions would be in the style of, “I noticed that you <action> can you explain why you did this…” or “What do you think the <symbol> indicates?”. The researcher and

the participant would use the phone to get to this point in the task to allow them to relate it to the actual instance of the issue in question. Feedback from this was then recorded on the same sheet next to the problem that had been highlighted during the observation stage.

8. The participant was then asked to complete the SUS questionnaire for their current phone and then for the phone used during the task.
9. The Technology Familiarity Questionnaire was then administered including the additional questions.
10. Finally, the participant was asked if they had any questions and if they would like to see any of the notes that were written.

6.6 Results and Analysis

The results are split between those produced by using the vis-UI-lise tool and those from the user testing. These are then brought together to analyse the performance of the vis-UI-lise tool. This process is summarised in Figure 6-6 earlier in this chapter and the various outputs and analysis are described in the following sections.

6.6.1 Summary of vis-UI-lise Results

The vis-UI-lise tool predicted 69 potential problems across the 14 task steps and a number of possible global problems that could occur across one or more of these steps. An overall summary of the PowerPoint™ based output is shown in Chapter 5 with Figures 5-16 to 5-19, however, a more detailed example is given here in Figures 6-9 to 6-13. This covers task step 2 from Table 6-2, which concerns PIN entry to unlock the phone. Of note is the step has to be further broken down to cover the specifics of how the device works. More specifically the sub-task concerns swiping upwards to reveal the PIN entry screen. The 'before' and 'after' of this step are shown as recorded by the vis-UI-lise tool in Figure 6-9. Looking at the upper half of this figure it shows that there is no visual indication of the need to enter a PIN or how to get to this point. The function is essentially 'missing' from a visual standpoint. This invisibility has been addressed to a degree in the UI by the fact that the time and date text bounces up and down slightly if the user presses the screen. However, if the user tries to swipe left or right, as opposed to a single direct press then there is no bounce. The lack of feedback for an attempted horizontal swipe is potentially significant as other devices, such as Apple iOS™, have previously had a right to left swipe to unlock (this has now been replaced by finger or face biometric authentication). Therefore, some users may naturally try this and will not get the bounce feedback. So even with the problem being mitigated to a degree, by the 'bounce', this step represents a clear example of invisibility. The full analysis of this task step for the 'function', 'operation' and 'feedback visibility is shown in Figures 6-10, 6-11 and 6-12. This leads to a set of predicted usability problems shown in Figure 6-13.

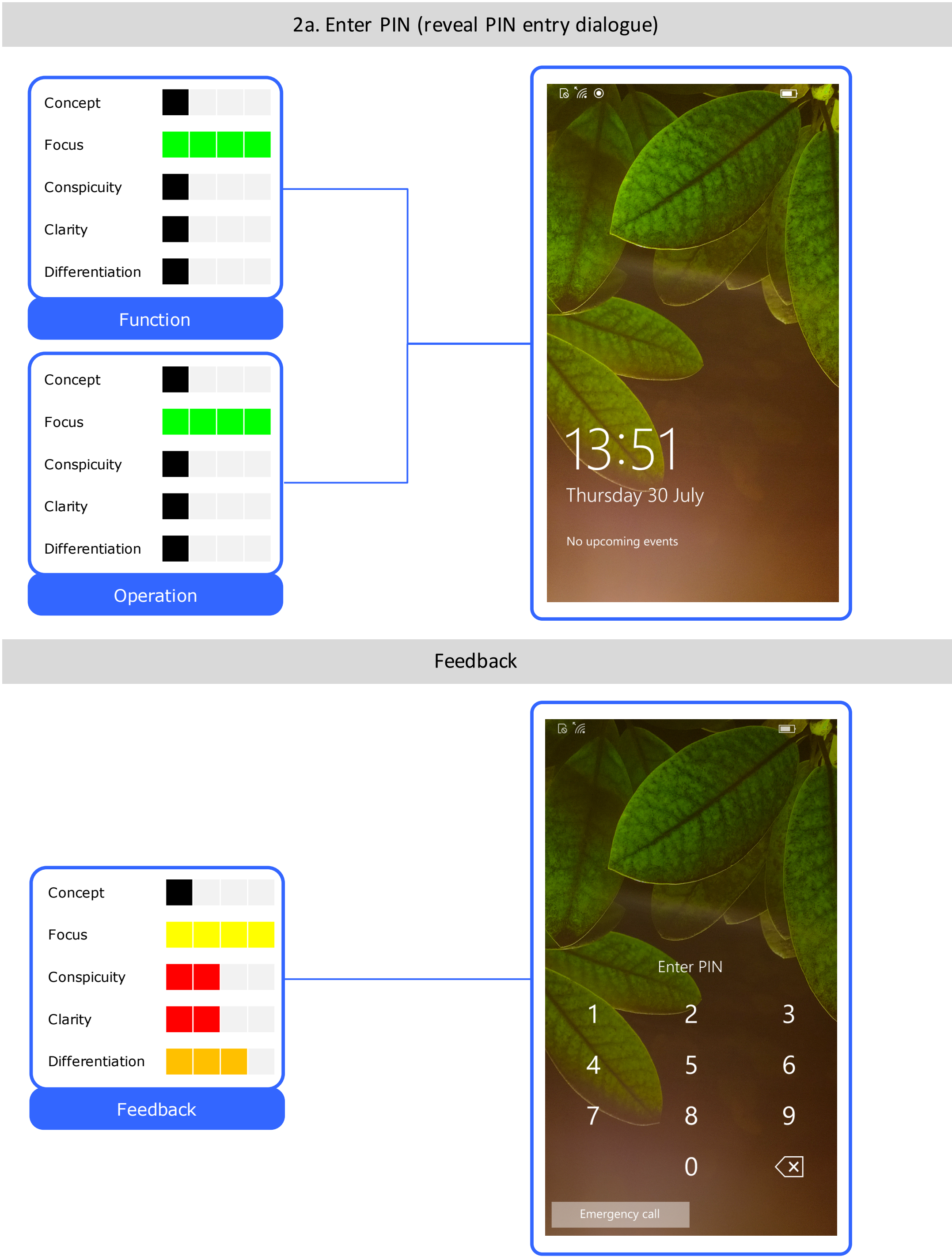


Figure 6-9: A copy of the results from using the vis-UI-lise tool for the PIN entry task step

FUNCTION ANALYSIS: 2a. Enter PIN (reveal PIN Entry Screen)				
Aspect	Question	Answer		Score
Concepts				
	What is the concept of the 'function'?	A slider from bottom		
	How is the concept conveyed visually?	Nothing explicit is shown on the lock screen		
	Is it a **familiar** concept to the user? (check against real examples)	Yes. Other devices have slide to unlock but with some form of indication and it is typically horizontal		
	How well is the concept **represented** & are there elements **missing**?	The are no visual indicators of the function		
	Are there general **variations** of this concept that could cause **confusion**?	Yes. Other devices have slide to unlock but with some form of indication and is typically horizontal		
Line of sight & focus				
	Does the user have to move to get line of sight?	No		
	Can the user focus on it? (bi or vari focal glasses)	Can move the arm to bring it into focus		
Conspicuity				
	Is it in the central visual field?	It is the bottom half of the screen which is large and in view		
	Is it where the user would expect it to be?	If familiar with slide to unlock		
	How many other related elements are there?	4 icons, date, time, upcoming event		
	Does it stands out against other elements/background?	It is invisible		
Clarity				
	What are the key distinguishing features?	**Size**	**Contrast**	
	- No features directly related to function	Invisible	Invisible	
Differentiation				
	How different is it from other elements visible at the same time?	Invisible		
	How different is it from other elements visible at other times?	Invisible		
	Could it be confused with commonly used graphics/symbols that indicate something different?	Invisible		

Figure 6-10: A copy of the results from using the vis-UI-lise tool for the 'function' visibility analysis of the PIN entry task step

OPERATION ANALYSIS: 2a. Enter PIN (reveal PIN Entry Screen)				
Aspect	Question	Answer		Score
Concepts				
	What is the concept of the 'operation'?	An invisible slider from bottom with only the bottom half of the screen enabling a full swipe (upper half active)		
	How is the concept conveyed visually?	Nothing explicit is shown on the lock screen		
	Is it a **familiar** concept to the user? (check against real examples)	Yes. Other devices have slide to unlock but with some form of indication and they are typically horizontal		
	How well is the concept **represented** & are there elements **missing**?	The are no visual indicators of the operation		
	Are there general **variations** of this concept that could cause **confusion**?	Yes. Other devices have slide to unlock but with some form of indication and is typically horizontal		
Line of sight & focus				
	Does the user have to move to get line of sight?	Assume already looking at it		
	Can the user focus on it? (bi or vari focal glasses)	Can move the arm to bring it into focus		
Conspicuity				
	Is it in the central visual field?	Assume already looking at it		
	Is it where the user would expect it to be?	Yes if identified as a slider		
	How many other related elements are there?	Date, time and upcoming events		
	Does it stands out against other elements/background?	It is invisible		
Clarity				
	What are the key distinguishing features?	**Size**	**Contrast**	
	- No features directly related to function	Invisible	Invisible	
Differentiation				
	How different is it from other elements visible at the same time?	Invisible		
	How different is it from other elements visible at other times?	Invisible		
	Could it be confused with commonly used graphics/symbols that indicate something different?	Invisible		

Figure 6-11: A copy of the results from using the vis-UI-lise tool for the 'operation' visibility analysis of the PIN entry task step

FEEDBACK ANALYSIS: 2a. Enter PIN (reveal PIN Entry Screen)				
Aspect	Question	Answer		Score
Concepts				
	What is the concept of the 'feedback'?	A moving date and time transitioning to a PIN pad		
	How is the concept conveyed visually?	Instantaneous movement of date and time with the swipe action		
	Is it a **familiar** concept to the user? (check against real examples)	Touch devices tend to have a slide or whole screen reveal		Red
	How well is the concept **represented** & are there elements **missing**?	The moving date and time is unusual and lacks contrast. When PIN entry is not required the whole screen slides up which is inconsistent		Red
	Are there general **variations** of this concept that could cause **confusion**?	Slide to unlock typically moves the whole screen or part of screen to reveal a keypad		Red
Line of sight & focus				
	Does the user have to move to get line of sight?	Assume already looking at it		Green
	Can the user focus on it? (bi or vari focal glasses)	Can move the arm to bring it into focus		Green
Conspicuity				
	Is it in the central visual field?	Assume already looking at it		Green
	Is it where the user would expect it to be?	Yes if identified as a slider		Green
	How many other related elements are there?	Depends on the background image		Red
	Does it stands out against other elements/background?	The date and time have poor contrast versus the background image		Red
Clarity				
	What are the key distinguishing features?	**Size**	**Contrast**	
	- Moving date and time	Large	Poor against background	Red
Differentiation				
	How different is it from other elements visible at the same time?	The date and time move and the PIN pad appears		Amber
	How different is it from other elements visible at other times?	The date and time moves		Amber
	Could it be confused with commonly used graphics/symbols that indicate something different?	A moving date and time is unusual to indicate a non time function		Amber

Figure 6-12: A copy of the results from using the vis-UI-lise tool for the 'feedback' visibility analysis of the PIN entry task step

Usability Problem Prediction

Task Step: 2a. Enter PIN (reveal PIN Entry Screen)				
Issue	Visibility problem	Probability	Impact	Rationale
The user fails to identify that there is a slide to unlock function completely	All	Low	High	The control is invisible so it is reliant on prior knowledge or trial and error. The bounce feature on pressing is limited to the date and time
The user fails to initially identify that there is a slide to unlock function	All	High	Low	The control is invisible so it is reliant on prior knowledge or trial and error
The user tries one of the buttons to unlock	All	Medium	High	The control is invisible so it is reliant on prior knowledge or trial and error
The screen times out while the user is trying to work out what to do	All	Low	Low	Quite a long duration time out and the user will probably try again
The users fails to perceive the movement of the date and time when correctly sliding	Clarity	Low	High	The concept is unusual combined with poor contrast and potential obscuration of date and time by the user's hand

Figure 6-13: A copy of the results from using the vis-UI-lise tool covering the predicted problems resulting from the analysis of the PIN entry task step

6.6.2 Summary of User Testing Results

The user testing followed the same 14 step task sequence as that used for the vis-UI-lise evaluation and produced the set of results set out in Table 6-3 as follows. This includes the results from the additional instruments aimed at characterising the users, namely the System Usability Scale (SUS) and the Technology Familiarity Questionnaire (TFQ) described in Section 6.5.4. As previously stated, the aim of these was to help provide a more nuanced understanding of the data set and also potentially guide future work.

Table 6-3: Individual user results

User	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Sex	M	M	F	F	F	F	M	M	F	M	F	M	F	M
Age	71	69	75	67	58	54	62	56	55	58	52	63	60	52
Problems	42	29	58	26	21	19	32	6	7[3]	9	30	23	18	29
SUS[1]	70	63	43[2]	63	23	30	60	73	13	38	65	28	75	83
TFQ	70	73[4]	48	64	56	52[5]	55	55	72	85[7]	65	67	53	72
Selfie	Y	N	N	Y	Y	Y	Y	Y	N[6]	Y	Y	N[6]	N	Y
Smartphone	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

1. Only the results from the 'test' phone were analysed and values were rounded up to aid readability
2. User modified questions to makes sense to her (see discussion section for more details on why this was allowed)
3. Video failed and problem numbers taken from notes
4. Partial TFQ completed and this is a normalised value based on the completed part
5. User ticked two adjacent columns as they wanted the value in between. The calculation was done by distributing the number evenly between both columns
6. They had seen Apple FaceTime™ used for making video calls where the user has to switch from the rear camera to the front-facing one
7. Did not complete frequency of use rating on the 'other category'

6.6.3 User Characterisation

Although the sample size is small some basic statistical analysis was performed to see if there are any correlations between the instruments and the number of problems that a user experienced. Scatter plots were examined as a starting point and then simple linear regression analysis was performed using Microsoft Excel (MacOS v16.25) to see if there was any correlation. The most pertinent scatter plots are shown in Figures 6-14 to 6-16 below. These have a linear regression line plotted by the Excel function. The number of problems that the user experienced represents the dependent variable and the SUS, TFQ and age the independent variables.

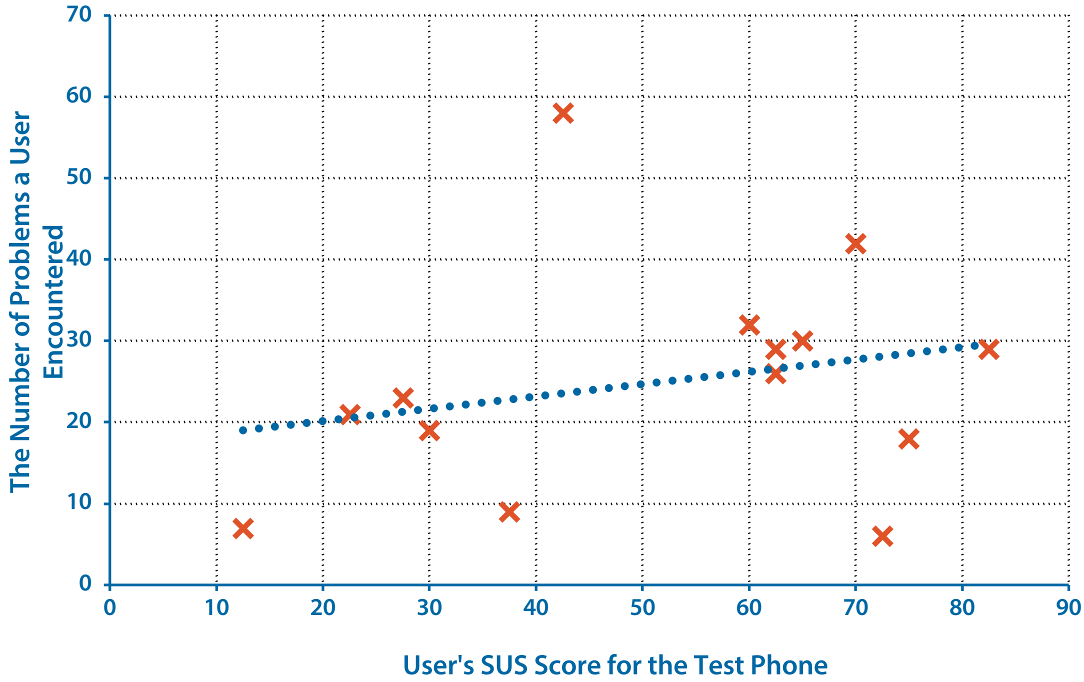


Figure 6-14: A scatter plot of the total number of problems each subject experience versus their TFQ score

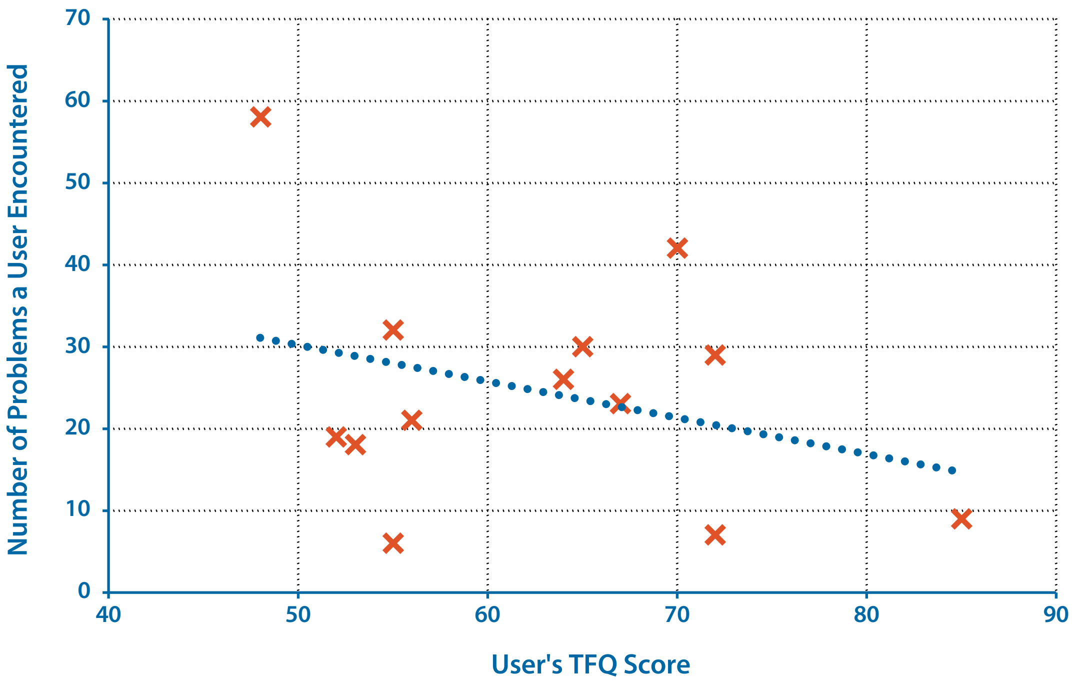


Figure 6-15: A scatter plot of the total number of problems each subject experienced versus their TFQ score

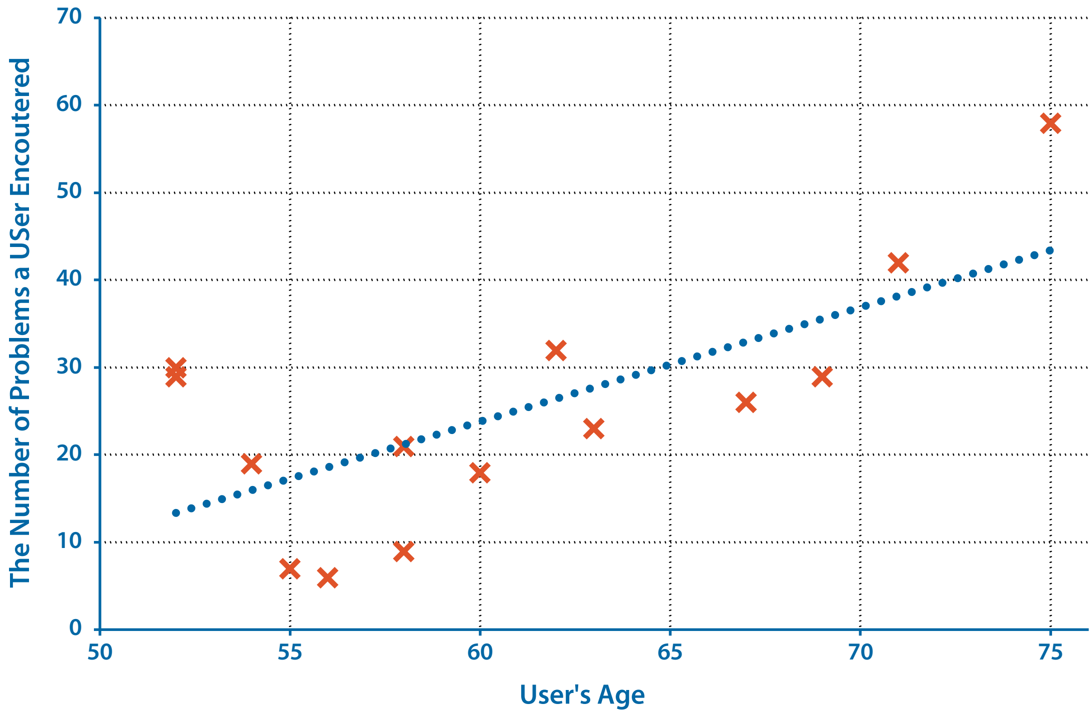


Figure 6-16: A scatter plot of the total number of problems each subject experienced versus their age

The results of this analysis are shown in Table 6-4 as follows.

Table 6-4: Linear regression analysis of user characteristics giving the Pearson correlation coefficient and associated p-value

Variables	No.	Pearson	p-value
SUS vs. Problems	14	0.24	0.40
TFQ vs. Problems	13[1]	-0.32	0.28
Age vs. Problems	14	0.68	0.007

1. The TFQ analysis is only for 13 participants as one of the participants had only partially completed the questionnaire. Analysis including the corrected value for this participant makes little difference to the correlation.

With regards to the SUS score, it is reasonable to expect a negative correlation as a high SUS score equates to greater confidence in the use of a product. In other words, the higher the SUS score the fewer problems you would expect a participant to experience (See Figure 6-14). However, in this case, there is a negligible positive correlation (Mukaka, 2012) and the p-value shows it is not statistically significant.

In the case of the TFQ, it is again reasonable to expect a negative correlation and it is negative with a negligible to low correlation (Mukaka, 2012), but the p-value shows it is not statistically significant. With 'age' one would expect a positive correlation, in other words, the older a person is the greater number of problems they would experience. Indeed, the analysis shows a correlation bordering on a high positive correlation (Mukaka, 2012) which is also statistically significant. This is not a surprise as age is well understood to have an impact on technology use (Czaja, 2005; Blaschke, Freddolino and Mullen, 2009; Bradley, Hosking, Langdon and Clarkson, 2017). However, correlation is not causation and of particular note is that the cohort bridges those in work and those now in retirement so other factors may be in play over and above the age-related issues such as cognitive decline.

The negligible correlation with the SUS and TFQ is disappointing as the value of being able to aid participant selection and characterise any results is significant. However, the sample size is small, and it is an area that warrants further work for usability testing in general but is beyond the scope of this work to do so. As it stands, the simple rule of thumb of using older users to find problems holds true in this case. If further testing was to be undertaken the selecting older users would make sense as a way of finding edge cases.

6.6.4 Characterisation of the Problem Set

The 14 participants resulted in the identification of 349 instances of a usability problem. When repeat problems are removed there were 113 different types of problems seen.

A plot of the accumulative number of new problems found for successive users in the test is shown in Figure 6-17. Curve fits were performed using a simple least-squares approximation regression analysis, with the accumulative binomial probability formula discussed in section 6.2.2. The first fit was performed to reach the final total of 113 and has a probability of 0.25. The second improved the curve fit by reducing the final value to reduce the least-squares result, this leads to a ‘final value’ of 104.9. Using this value puts the probability at 0.30. This highlights the issues raised on earlier regarding the binomial formula. Here the last two subjects uncover more problems than would be predicted if the testing had stopped at 12 subjects. This is indicative of the likelihood that the probability of a problem occurring is heterogeneous and not homogenous.

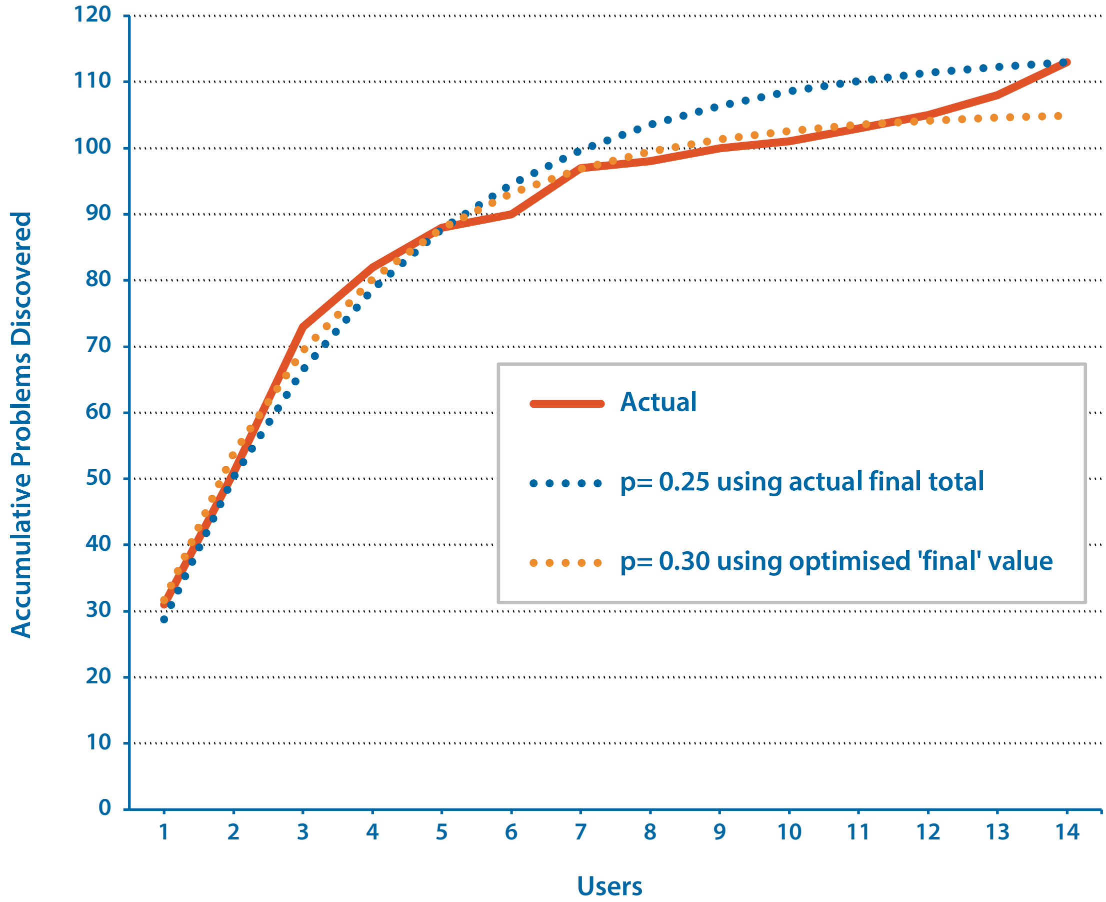


Figure 6-17: A plot of the accumulative total of new problems found for each successive user. Curve fits for different probabilities are shown.

As it stands a probability of 0.25 in the binomial model suggests that 98% of problems have been found and a further 2 problems remain to be found. As has been stressed this is predicated on the discovery rate being homogenous which it is highly unlikely to be. This probability is comparable to ranges from other studies discussed earlier in Section 6.2.2 which are 0.12 to 0.58 (Nielsen and Landauer, 1993) and 0.16 to 0.42 (Hartson, Andre and Williges, 2003). It puts this result and the further optimised value in the middle of these ranges. This variability in the probability of a problem occurring means it is impossible to say how big the set might be as there could be numerous problems that are highly unlikely to occur that were not observed. However, this analysis does help characterise the problem set as being comprehensive but clearly not exhaustive. Therefore, it fulfils its purpose of being a benchmark for comparison, albeit with appropriate caveats. Further issues around this will be elaborated on in the discussion section.

6.6.5 Simple Comparison of Predicted versus Actual

A basic list of predicted and actual problems is provided to provide a simple view of the magnitude of the problems considered and an overview of the predictions that came true in testing. Table 6-5 contains the predicted problems. Each prediction has a reference ID that starts with the task number and a simple incrementing number for the problems associated with the task. When the predicted problem was seen in practice the background of the table cell is coloured green to indicate this. In the list of problems that did occur (Table 6-6) those that were predicted are again highlighted in green and labelled with the prediction reference from Table 6-5.

Table 6-5 shows that a single prediction can have a one-to-many relationship with different observed problems. This is because a predicted problem can be sufficiently broad to encompasses slightly different examples of that problem. For example, the first prediction was that the user would not find the power button. In practice, users failed to do this in several different ways. These were recorded separately, as there may be subtle issues that are worthy of further investigation. For example, one user pressed the phone's logo instead of the power button. This was recorded separately from other forms of failure to press the power button. It can be argued that the prediction, such as the one around the power button, was too imprecise. However, elucidating all the possible ways that a user may try to turn the device on may be of limited use and doing this across all the predicted problems would be onerous. Table 6-6 shows that there are a significant number of one-to-many relationships. This further highlights the difficulties with usability problem description which will be discussed later.

Table 6-5: Simple list of usability problem predictions with the ones that occurred in user testing marked in green

Task Step: 1. Turn phone on	
1.1	Not finding the power+lock button at all
1.2	Only finding the power+lock button by trial and error e.g. pressing the camera or volume keys N.B. pressing camera key will launch the camera and potentially cause a whole sequence of problems
1.3	Not pressing the button for long enough leading to rejection that it is the correct button
1.4	Confusion between lock & power state i.e. the user may press for too long and go into power off sequence
Task Step: 2a. Enter PIN (reveal PIN Entry Screen)	
2a.1	The user fails to identify that there is a slide to unlock function completely
2a.2	The user fails to initially identify that there is a slide to unlock function
2a.3	The user tries one of the buttons to unlock
2a.4	The screen times out while the user is trying to work out what to do
2a.5	The user fails to perceive the movement of the date and time when correctly sliding
Task Step: 2b. Enter PIN (input numbers)	
2b.1	The user fails to perceive the buttons as buttons
2b.2	The user enters the wrong or too many numbers
2b.3	The user enters the wrong or too many numbers without knowing it
Task Step: 2c	
2c.1	The user may not read the "Incorrect PIN" message in time and be confused about what has happened
2c.2	The user may not count the number of incorrect entries and put the phone into challenge mode
2c.3	The user may not understand the challenge mode and the need to press the input field to enter the challenge phrase
Task Step: 2d. Enter PIN (check PIN)	
2d	This is appropriate automation that should not lead to direct problems but may cause problems in building up mental models of when confirmation is manual e.g. SIM PIN entry which is not automatic
Task Step: 3. Start camera	
3.1	The user may not identify the camera app at all
3.2	The user may select another app e.g. photos
3.3	The user may attempt to find a dedicated camera button and try the other silver buttons (volume & lock)
Task Step: 4. Switch to front facing camera	
4.1	User fails to identify switch camera button
4.2	User tries other buttons first
4.3	User gets confused about which camera is in operation and moves phone around to try and determine which it is
4.4	The user may be confused about the fact the phone has a front facing camera
Task Step: 5. Compose Picture	
5.1	The user may remain confused over which camera they have selected
5.2	The target image(s) may be out of the field of view and the user may have a problem aligning the camera as the target is not in their field of view as it is with conventional photography
5.3	The user may block the camera not knowing where it is
Task Step: 6. Take Picture	
6.1	The user may fail to find the shutter button and give up
6.2	The user may press another button thinking it is the shutter button
6.3	The user may fail to realise the picture has been taken
Task Step: 7. View picture	
7.1	The user fails to identify the photo app button and gives up
7.2	The user tries other buttons and becomes confused
7.3	The user presses the start (home) key to get to the photos app (valid but inefficient route)

Task Step: 8a. Create and email with picture – display share options	
8a.1	The user may not try anything because nothing is explicitly 'share'
8a.2	User may not expect a list and the icon does not indicate that it is a menu and may become confused and explore other options e.g. back or home button
8a.3	The user may not correctly understand that it is the photo library app
8a.4	The user may be confused by the disappearing header and footer. Leading to exploring incorrect options such as using the start (home) key
8a.5	The user may press the "View Collection" button thinking it contains the recently taken image
Task Step: 8b. Create and email with picture – create prepopulated email	
8b.1	The user may not understand that the screen is a menu
8b.2	The user may become confused between messaging and email
8b.3	The splash screen may cause confusion leading to the user thinking they have done something wrong leading to them taking an incorrect action e.g. selecting start (home)
Task Step: 9. Enter email address	
9.1	Input field not identified leading to pressing other part of message
9.2	Unable to find numeric input panel
9.3	May struggle to understand 2 available forms of cursor navigation required to edit
9.4	Auto completion pop-up may cause confusion leading to incorrect actions e.g. "Use this address: ….." comes up before the address is complete
9.5	Mistyping
Task Step: 10. Enter email title	
10.1	Email title Input field not identified as such leading to pressing other part of message
10.2	May struggle to understand 2 forms of cursor navigation available for editing
10.3	Mistyping
10.4	May be confused by suggested words above keyboard and their function
Task Step: 11. Enter email message	
11.1	Attachment area title looks like a placeholder for message text and may cause user to press it to enter message
11.2	The cursor mode may cause confusion and lead to incorrect actions
11.3	Moving the cursor to the required place for text editing may be misunderstood
11.4	May be confused by suggested words above keyboard and their function
Task Step: 12. Send	
12.1	Not identifying the button and giving up
12.2	Pressing the wrong on screen button
12.3	Pressing one of the navigation buttons
12.4	Task Step: 13. Return to Home Screen
12.5	The user fails to identify the home button and gives up
12.6	The user mistakenly chooses another button
Task Step: 13. Return to Home Screen	
13.1	The user fails to identify the home button and gives up
13.2	The user mistakenly chooses another button
Task Step: 14. Put into standby	
14.1	The user presses the wrong button e.g. camera or volume
14.2	The user presses the button for too long and powers the phone off
14.3	The user looks for a software based standby mode instead of a button
Relevant global functions: Navigation Keys	
GA1	Confusion over where 'back' goes and ending up moving out the required app
GA2	Accidental activation of app switcher (from back)
GA3	Accidental activation of one hand mode (from home)
GA4	Confusion over nature of Cortana (from search)
Relevant global functions: App Launcher	
GB1	The user gets confused about the nature of the difference between the home screen and "All apps"

Table 6-6: List of unique problems that occurred during user testing. The ones predicted are marked in green (as opposed to red) with reference to the problems predicted in Table 6-5.

Task Step: 1. Turn phone on	
1.1	Presses top right of screen (Nokia Logo)
1.1	Presses top left of screen
1.1	Presses Back Button
1.1	Presses Glance Screen Text
1.1	Presses Speaker
1.2	Swiped phone screen to unlock
1.2	Pressed volume down screen to unlock
1.2	Appears to attempt press of camera button while phone is waking up after correct press of power button
1.2	Presses home 'button'
1.2	Taps screen
1.2	Pressed search key
1.2	Presses front facing camera
1.3	Insufficient press of on key
1.4	Pressed power button for too long to put into power down mode
Task Step: 2a. Enter PIN (reveal PIN Entry Screen)	
2a.2	The users swipes from left to right like an iPhone but at a slight angle and then left to right
2a.2	Taps screen
2a.2	Presses home 'button'
2a.2	Presses search button
2a.3	Presses back button
2a.3	Presses volume down button
2a.3	Presses power button (puts phone to sleep)
2a.3	Presses camera button
2a.2	Presses search button
2a.2	Presses home button
2a.2	Presses status icons at the top left
GA3	Presses and hold home button and puts it into one handed mode (deemed unique as the first time it has occurred at this task step)
2a.2	Swipes from top to bottom (first top to bottom swipe)
GA3	Pressed and hold home button into one handed mode (first instance at this step)
2a.5	User swiped up but failed to recognises the sliding
	After correctly swiping up immediately swipes down
	User double pressed search button (appears to due to rapidity of presses
	User double pressed back button
	User double pressed home button
Task Step: 2b. Enter PIN (input numbers)	
	Presses delete button rather than entering PIN. This it could be argued is a problem with not reading the instructions see below.
Task Step: 2c	
Task Step: 2d. Enter PIN (check PIN)	
2d.1	User appears to move finger to press non-existent confirmation button
Task Step: 3. Start camera	
GA4	Scrolls through information presented about from global search
3.2	Pressed photos app instead of camera
	User fails to make press of Phone icon register
Task Step: 4. Switch to front facing camera	
4.2	Tapped shutter button and takes picture (implication that trying to switch cameras to front facing)
4.2	Presses settings button
4.2	Selects Lumia Selfie which goes to PIN entry (this is due to the fact that camera is partially operable while locked)
4.2	Presses exposure options (linked problem where failed with the correct option leading to exploring others)
4.2	Presses shutter release button apparently to switch cameras
GA3	Puts in one handed mode (first time this has occurred at step 4)
	Unsuccessful attempt to press switch button
	Repeated attempt
Task Step: 5. Compose Picture	
5.2	User struggled to align camera for selfie
Task Step: 6. Take Picture	
6.3	Taps screen after taking picture
6.3	Presses flash setting button
	Taps screen
Task Step: 7. View picture	
7.2	Presses exposure menu 'expand' hoping it is for save pictures
7.2	Presses '…'
7.2	Presses back key
7.2	Presses camera button (first instance at step 7)
7.3	Hovered over different options and pressed home key which went to lock screen
	Attempted press of home button not registered
	Failure to select Photos amongst icons resulting in scrolling
	Accidental activation of "Films & TV" app while scrolling
Task Step: 8a. Create and email with picture – display share options	
8a.4	The bottom menu disappears while the user is viewing it
8a.4	Option menu disappears and activate zoom function
8a.4	Pressed home key (unique to this instance)
GA2	Press and hold back to put into task switcher (first time at this step)
	Presses further options button (3 dots)
	Presses photo editor option
Task Step: 8b. Create and email with picture – create prepopulated email	
8b.1	Pressed share menu title
	Selects "tap to share" which is for Bluetooth sharing
Task Step: 9. Enter email address	
9.5	Types a "&" instead of "@" difficult to say why
	Sends emails without changing title or entering message
	Deleted incorrect text "Hello" in address field and correct address
	User confused by semi-colon after email address is entered
Task Step: 10. Enter email title	
10.1	Fails to correctly select the title field which results in "hello" being entered in address field
10.1	Selects body of message instead of title field
10.3	Mistypes
10.1	Pressess 'dead area' between keyboard and button bar
10.1	Presses input microphone
Task Step: 11. Enter email message	
11.3	The 'dual cursor' display means the user does not tap the text field so the text entry for the body of the email goes into the Title
	Could initially work out how to scroll across on predictive text to see partially displayed word. Works this out later.
	Tried to select "Department" from suggested list but accidentally pressed "r" key
	User did not put in a carriage return. This may have simply been that they did not realise that one was required, particularly with the informal tone of the email.
Task Step: 12. Send	
12.1	User struggled to 'see' the send button
12.3	Presses home button & not clear why
	Incomplete email addresses means sending not possible
	Tried sending with incomplete email address
Task Step: 13. Return to Home Screen	
13.2	The user tapped the centre of the screen
13.2	The user presses the volume up or down button
13.2	The user presses the power/standby button and puts phone into standby mode
13.2	The user presses the camera button and activates camera
13.2	User is confused by camera button and states that it takes you back to where you were
13.2	Presses camera menu
GA3	Presses and holds home key an activates one hand mode
GA2	Presses and holds back key and activates task switcher
GA4	Presses search
GA4	Tries to use Cortana to get help but doesn't press microphone button. The screen says "ask me anything"
GA4	The phone then automatically brings up a daily summary of items by default that coincides with speaking thus potentially confusing the user that it has "listened" to what was said
GA4	On a second attempt the user does press the mic button the question the user ask is misinterpreted as a request for directions presumably due to the use of the word home.
13.2	User pressed sync button (deemed to be unique example of pressing one of many wrong options identified in 13.2)
13.2	User presses setting button (deemed to be unique example of pressing one of many wrong options identified in 13.2)
13.2	User presses delete but photo not selected (deemed to be unique example of pressing one of many wrong options identified in 13.2)
	Takes screen shot by simultaneously pressing the power and volume up buttons
	Swipe left at bottom of screen
	Tries swiping to get out of camera
Task Step: 14. Put into standby	
14.1	User presses camera button (starts camera app)
14.1	User presses camera button again (takes picture)
14.2	Pressed power key for too long putting it into shutdown mode
14.3	User tries to press back button but fails to register. Effectively a stacked double problem (similar to A35 but without hold therefore deemed to be unique)
14.3	User tries to press search button but fails to register Effectively a stacked double problem (similar to A36 but without hold therefore deemed to be unique)
14.3	Successfully presses back key but incorrect option (similar to A35 and 128 but deemed a significantly distinct variant as leads to different outcome)
14.3	Scrolled up screen looking for options
14.3	User presses home button (first instance of using home button)
14.3	Scroll and presses settings 'app'
	Does not press power button for long enough
	User went to top right like an iPhone to turn off where there is no button

6.6.6 Performance Calculations

With all the problems associated with the binomial model and usability problem set size, it is simpler to consider the performance measurements in comparison to those actually found i.e. with a certain number of test subjects, in this case, 14. As the binomial model here is only predicting a further couple of unfound problems, this means there is little difference in practice anyway. This means that rather than stating the performance regarding a theoretical set size it is stated against the actual number of problems found for a specific number of test subjects. This is transparent and also useful in real-world terms of one can see the value of the evaluation tool in comparison to user testing.

Thoroughness

Thoroughness is calculated as follows for all the problems found i.e. including duplicates and only for unique problems. The equation is modified to reflect prediction versus actual and actual is restricted to problems seen as opposed to the potential problem set as discussed in section 6.4 above.

$$Thoroughness = \frac{Number\ of\ real\ predicted}{Number\ of\ real\ problems\ found}$$

Table 6-7: Thoroughness results

Problem set	Calculation	Result	
All problems found	311/349	0.89	89%
Unique problems	85/113	0.75	75%

The results in Table 6-7 and subsequent calculations are quoted in terms of a simple calculated value in line with Hartson et al. (2003) and as a percentage as this is deemed to be a helpful representation of what the calculation is showing.

Validity

Validity is the proportion of problems predicted by the candidate UEM that are real usability problems.

$$Validity = \frac{Number\ of\ real\ problems\ found}{Number\ of\ issues\ identified\ as\ problems}$$

Table 6-8: Validity results

Probability	Calculation	Result	
All	27/66[1]	0.41	41%

1. Numbers are from Table 6-5. N.B. many of the 27 problems that 'occurred' in testing have a one-to-many relationship with observed problems as highlighted in Table 6-6.

Table 6-9: Validity results excluding special cases

Predicted probability	Occurred	Did not occur
Very low	7	2
Low	5	20
Medium	6	13
High	8	3
Special cases	1	1

The special cases are predicted problems that were deemed to be outside the likely bounds of what might occur. The first concerned the phone entering a 'security challenge mode' if a number of failed PIN entries were made. The second concerned the potential confusion between PIN entry for the devices as opposed to unlocking the SIM (which occurs when a device has been fully powered off or rebooted). PIN entry for the device has an automated 'ok' i.e. when the required number of digits is reached is automatically checks. However, with SIM PIN entry there is an okay button (this is for technical reasons in that the phone does not know the length of the SIM PIN). For the first special case, this did not occur, as no one made multiple failed attempts. However, interestingly the second case did occur where a user appeared to move their finger towards a non-existent 'ok' button. These two examples highlight how intricate problems are and also that predicting probability is hard as the latter case was deemed to be outside the scope of the scenario, yet it still occurred. Despite these specific examples, the validity result has to be seen in light of the fact that low probability problems are unlikely to be seen with the sample size of 14.

Effectiveness

The effectiveness is calculated using the values from the previous two sections as follows:

$$Effectiveness = Throughness \times Validity$$

Table 6-10: Effectiveness Figure of Merit

Problem set	Calculation	Result
Unique Problems	0.75 x 0.41	0.31

The effectiveness score is impacted by the issues for the validity number described above i.e. not seeing low probability problems in testing.

Reliability

As previously discussed (e.g. section 3.11), reliability concerns the reliability of evaluation results across different evaluators. This vis-UI-lise evaluation was performed by one person and therefore it is not possible to assess this directly. However, it will be indirectly addressed as part of the evaluation with practitioners which is covered in the next chapter.

Downstream Utility

Downstream utility concerns the ability to positively improve the usability of a user interface. As mentioned previously from the work of Blandford et al. (2008) is the notion of 'persuasive power'. Considerable effort was put into the format of the vis-UI-lise tool. In particular the use of 'before' and 'after' images of the user interface and the tabular format of results with traffic light format. The aim of this is to make it very quick to identify the screen in question and the issues that have been identified. This is something that will be picked up again in the next chapter with feedback from practitioners.

Cost-Effectiveness

There are two key points of note with regards to cost-effectiveness. The first is that the vis-UI-lise tool is well suited to early-stage evaluation as it can be performed on 'paper' prototypes. In other words, it does not require a functioning system, although the functionality would need to be described. This enables problems to be found and corrected earlier which can avoid costlier downstream changes that can be higher by a factor of 50 to 200 (Boehm and Papaccio, 1988; Boehm, 2006). The tool could easily be used to assess proposed changes too.

Secondly, the tool avoids the cost and time of testing with real users. This is why the 'thoroughness' and 'validity' calculations are important to understand the number of problems that may be missed and the number that may not exist in practice. The initial results on this are encouraging and will be discussed further in the next section. However, it is reasonable to propose that the vis-UI-lise tool could be used in addition to user testing but could be used earlier and would allow any user testing to be focused on areas of concern or uncertainty. Therefore, overall it could lead to a reduction in user testing and therefore a cost-saving.

Visibility

In addition to the established metrics for UEM evaluation, it is worth considering the proportion of problems predicted and identified that can be attributed to being visibility issues. Table 6-11 outlines the proportion of problems that are visibility related for both the predictions from the vis-UI-lise tool and what was observed. This is further broken down in terms of all the problems seen and those that are unique problems. Not unsurprisingly the vis-UI-lise tool produces almost exclusively visibility-based problems as this is the focus of the tool. However, of note is that it did help identify 3 non-visibility problems. Perhaps more surprising is the high proportion of actual problems observed that could be attributed to visibility issues, in the case of unique problems this was 81%.

Table 6-11: Proportion of problems that are visibility related

Problem set	Calculation	Result
vis-UI-lise prediction	63/66	95%
All problems observed	298/349	85%
Unique problems observed	92/113	81%

6.7 Discussion

Before discussing the various performance measures, it is important to briefly discuss the issues encountered with usability problem description. This is consistent with the issues outlined in the literature and discussed in Section 3.8 in Chapter 3.

6.7.1 The problems with problem description

One of the known problems in usability evaluation and the evaluation of usability evaluation methods is defining the boundaries and nature of a problem which was discussed in Section 6.3. Not unsurprisingly this proved to be a problem in this research too.

A deeper look into what a usability problem is becomes an issue of causality that is an area of intense philosophical debate over the course of human history (Schaffer, 2016; Menzies, 2017). It is therefore not a surprise that the issues around it are hard to resolve as causality is inherently complex. In particular, it is hard to put a boundary around a problem—where the problem starts and ends. A simple example is activating the camera on the phone that was tested. A screenshot from the task sequence is shown in Figure 6-18 as follows.

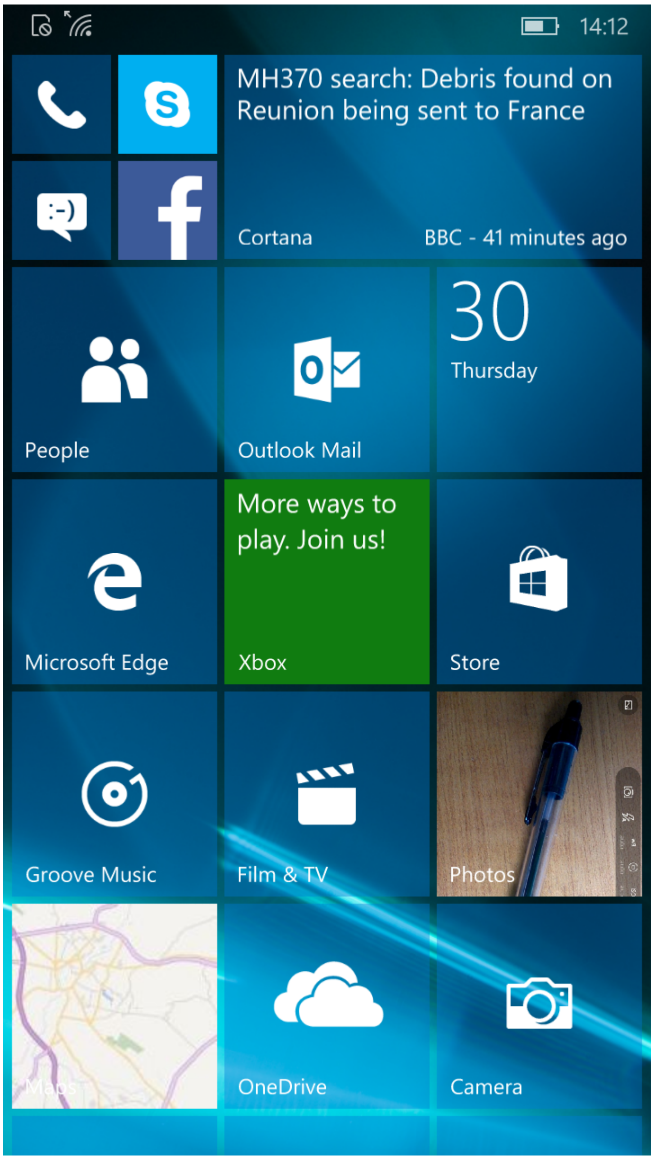


Figure 6-18: Screenshot from Task Step 3 showing the 'Photos' and 'Camera' app buttons at the bottom right of the screen

At first sight, the situation looks clear. There is a button that has a camera icon and the text 'Camera' on it. With a narrow system boundary, this looks like a good design with a combination of an icon and text label. Broadening the boundary further the 'Photos' app is directly above. The functionality of 'picture taking' and 'picture viewing' is separated, introducing potential confusion. The 'Photos' app does show the last picture taken helping to make the distinction. However, broadening the boundary further still shows that there is a dedicated camera button on the side of the phone (Microsoft, 2014) that can be used to both launch the 'Camera' app and doubles up as a camera 'release' button. Also, the 'Photos' app can be accessed from within the 'Camera' app. Also, when the phone was launched there were two different camera apps the default Microsoft one with the operating systems and an additional Nokia app. Finally, it is possible to rearrange the app button locations and the 'Camera' or 'Photos' buttons could be placed apart and off the initial screen area requiring the user to scroll down to see it. This then would mean that the user would not see the 'Camera' or 'Photos' enabling them to work out the difference between them.

The complexity around the camera could easily lead to problems being described from a simple description, that might cover multiple different 'failure' routes, to explicit descriptions for each problem. Therefore, any UEM tool and the evaluation of it is going to be hampered by the reality of this potential variability. Any results should be seen in this light.

6.7.2 What the measures show

The headline figures from the results are that the vis-UI-lise tool highlighted 66 potential problems. The user testing resulted in 113 unique problems being observed. As some of the predictions mapped to more than one observed problem, this resulted in 85 out of the 113 being 'predicted' in other words 75% (thoroughness). In terms of validity 27 of the 66 problems highlighted by vis-UI-lise were seen in the user testing giving a validity score of 41%. As was mentioned this figure needs to be seen in the context of the problem discovery rate i.e. it is unclear how many low probability problems there are that were not observed.

Having established the overall scores, it is useful to view these in comparison to the evaluation of other UEM's as described in section 6.4 and summarised in Figure 6-6. Although the performance measures calculated for the vis-UI-lise tool were based on the problem set of 14 users, as discussed, the problem discovery rate modelling (Section 6.6.4) suggests that it is a comprehensive set and likely to be comparable to other studies.

Hartson, Andre & Williges (2003) provide a helpful review of 19 papers from which they summarise the thoroughness and validity scores for various UEM's. Although such a summary is appealing it does decontextualize the numbers in terms of the system under evaluation and the specific elements of it that are being considered. It also highlights the variation in the number and type of evaluators involved in each study which is typically a team. As previously mentioned, 'the evaluator effect' (Hertzum and Jacobsen, 2003b; Capra, 2007; Hertzum, Molich and Jacobsen, 2014) means that any results have to be treated with caution. Therefore, an appropriate view is to see how the results from this research fit within the broad sweep of other results as opposed to trying to draw specific comparative conclusions such as whether it is better than a specific technique such as heuristic evaluation.

On this basis, a simple review of Hartson, Andre & Williges (2003) summary shows a range of values for 'thoroughness' from 15% to 90%. This excludes two outliers that are both less than 1%. For validity the range is 5% to 73%, again excluding the outliers. More recent studies such as Capra (2007) give a validity measure of 22% for two different cohorts of evaluators (students versus practitioners). This again highlights the evaluator effect as well as the variation in results. Finally, a study by Koutsabasis, Spyrou, and Darzentas (2007) compared a range of UEMs applied to a web-based system administrative system. What is useful about this particular study is the range of different evaluation approaches that were tested, albeit with the caveat that the evaluators were students. However, the previous study mentioned shows that students can perform well compared to experienced practitioners. In the study teams of 3 were assigned to one of four different evaluation approaches. These were: heuristic evaluation, cognitive walkthrough, think-aloud protocol, and co-discovery learning. Apart from the 'co-discovery learning' method, each approach was tested with 2 or 3 teams. Table 6-12 below collates these results from different tables within the paper.

Table 6-12: Results for different UEM's applied to a web-based administration system. Combining Tables 2,3, & 4 from Koutsabasis, Spyrou, & Darzentas (2007).

EMs	Thoroughness (%)	Validity (%)	Effectiveness[1]
HE1	24.3%	94.4%	0.23
HE2	24.3%	60.7%	0.15
HE3	20.0%	100.0%	0.20
CW1	25.7%	85.7%	0.22
CW2	24.3%	70.8%	0.17
T-AP1	27.1%	90.5%	0.25
T-AP2	24.3%	94.4%	0.23
T-AP3	20.0%	82.4%	0.17
C-D1	41.4%	74.4%	0.31

1. The Effectiveness scores are expressed as a number not a percentage in line with Hartson, Andre, and Williges (2003) definition.

These results show a couple of interesting things. Firstly, it reinforces the variability across UEMs and evaluators. Secondly, it shows a combination of low thoroughness but a high validity which is the opposite to the results using the vis-UI-lise tool. Again, highlighting how these results vary greatly and with it being reasonable to assume that strong contextual factors are driving this.

So, with the caution discussed above, it is reasonable to say that a 'thoroughness' score of 75% is very encouraging. The 'validity' score of 41% is harder to interpret due to the discovery rate problem but even with this issue, the overall 'effectiveness' figure of merit of 0.31 is again very encouraging. Considering that this the first full iteration of the tool, the results are certainly positive enough to warrant testing the vis-UI-lise tool with practitioners, which is ultimately the goal at this stage in the process. This testing is the subject of the next chapter.

6.8 Conclusion

Returning to the research questions that were addressed through this phase of the research.

IRQ3: *What can be done to improve it?*

> **SRQ 3a:** *How does a cognitive-based framework for understanding and representing the visibility of user interfaces compare to one based on a simple visibility threshold?*
>
> **SRQ 3b:** *How well can such a framework be embodied in a tool that predicts more usability problems than current approaches?*
>
> **SRQ 3c***: Can and will practitioners use such a tool to reliably improve the visibility of user interfaces?*

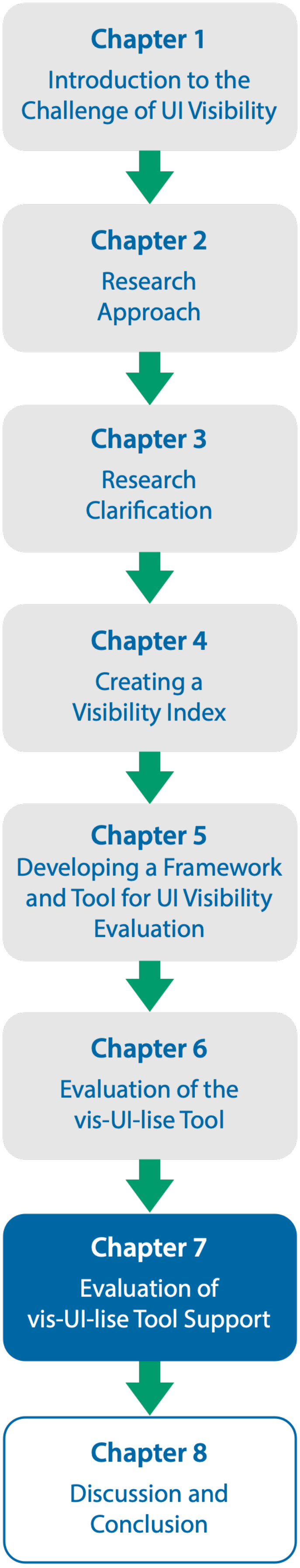
In addition, a useful by-product of this work was to gain further insights into the question:

> **SRQ 2b:** *Are these problems significant? (i.e. visibility problems)*

Addressing SRQ 2b first, the analysis of the unique problems observed showed that 81% of them (92 from 113) had a visibility dimension to them. This further strengthens the arguments made already for the importance of visibility in driving usability.

Moving onto IRQ3. With regards to SRQ 3a, answering this question is hampered by deeply rooted problems in evaluating UEMs articulated in the literature and experienced in this research. However, the pragmatic stance taken here is to compare the performance of vis-UI-lise to a broad range of results from the evaluation of other UEMs. The analysis of UEMs undertaken in Chapter 3 (see Table 3-10 for a summary) regarding how they address visibility showed that current UEMs do not take a cognitive-based approach. In other words, they are effectively, by default, taking a simple visibility threshold approach. Therefore, this comparison can be seen at a high-level as addressing this question. As such, the headline results (Thoroughness, 75%; Validity, 41%; Effectiveness, 0.31) are very encouraging and sufficient to warrant testing with practitioners.

With regards to SRQ 3b, the testing was undertaken with a ‘practitioner ready’ version of the tool and therefore again the results are positive with regards to this question. In other words, the framework can be embodied in a tool that compares favourably to current UEMs. This leads onto question SRQ 3c and how it works in the ‘hands’ of practitioners in the real world and this is the focus of the next chapter and represents the transition to Descriptive Study II of the DRM framework.


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 7

Evaluation of vis-UI-lise Tool Support

7.1 Introduction

Having established the underlying performance potential of the vis-UI-lise tool in the previous chapter the next step is to see how well it performs in the hands of practitioners. In DRM terms (Blessing and Chakrabarti, 2009, p.17) this represents the transition from the Prescriptive Study (PS) to the second Descriptive Study (DSII) which is shown by the arrows in Figure 7-1 below.

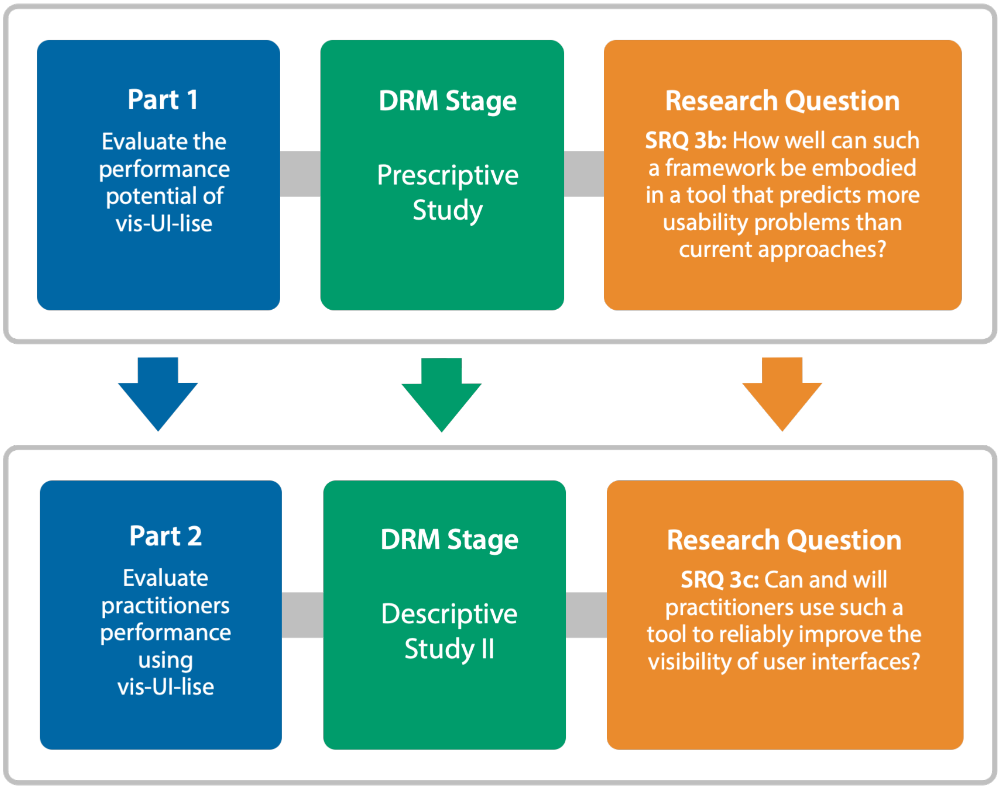


Figure 7-1: The transition to the Descriptive Study II which addresses the final supplementary research question

This moves the focus on to the final supplementary research question, SRQ 3c, as follows:

> **IRQ3:** *What can be done to improve it?*
>
> > **SRQ 3c:** *Can and will practitioners use such a tool to reliably improve the visibility of user interfaces?*

This chapter concerns evaluating the vis-UI-lise tool support and the wider context of how practitioners evaluate usability.

7.2 Tool Support Development

In Chapter 5 (section 5.5) the generic requirements for any usability evaluation method were discussed and are summarised as follows:

- **Scope & Positioning** [Practicalities]: What the proposed UEM will cover and how it is positioned concerning the development process and other evaluation approaches.
- **Output** [Insights derived]: The specifics of the output e.g. quantitative and qualitative aspects
- **Validity**
    - **Internal** [Reliability]: Internal validity concerns the consistency across different evaluators (inter-rater reliability)
    - **External**: External Validity addresses how well any results relate to 'real-world' performance
- **Usability**:
    - Productivity (effectiveness & efficiency)
    - Learnability
- **Persuasive Power**: Ultimately evaluation is only of value if it can effect change. Therefore, the approach and its output's ability to persuade practitioners to make appropriate changes is critical.

With regards to 'scope and positioning' and 'output', these were set in the development of the vis-UI-lise tool as described in Chapter 5. The 'validity' was addressed, in part, with the evaluation of the tool in Chapter 6. In this chapter the focus moves primarily to 'usability' and 'persuasive power' but in doing so leads to reflection of the 'scope and positioning' and how the 'output' addresses the needs of practitioners.

The details of the core vis-UI-lise tool were covered in Chapter 5. However, for the evaluation with practitioners, an associated support package was developed. The perspective taken was that there is a potential issue between the 'presentation' and actual 'function' of the tool. In other words, practitioners would be negatively influenced by something that looked experimental. Indeed, ultimately the tool would be compared to established tools/approaches. This relates to the issues of 'persuasive power' as discussed. Therefore, the decision was taken to develop a high-quality support package comprising of the following components outlined in Table 7-1 below.

Table 7-1: Elements of the vis-UI-lise support package

Item	Format
vis-UI-lise name & brand identity	Name & Logo
Training Presentation	MS PowerPoint
Evaluator Guide	A4 Booklet (PDF)
Quickstart Guide	Double-sided A3 Sheet
Practice Exercise	MS PowerPoint
Phone Example	MS PowerPoint
Blank Evaluation Template	MS PowerPoint

7.2.1 Branding

Discussion of branding may seem out of place in a piece of academic research. However, this research under the framework of the DRM is an iterative approach to developing a tool with the aim of implementation. To evaluate requires the 'attention' of practitioners for this evaluation and ultimately in use, once fully developed. It can be argued that this may unfairly influence the outcome, but as it has already been stated the comparison is with established tools, not with other experimental ones. Therefore, the aim is to make it comparable. Also, beyond simple branding, it helps in the explanation of the nature of the tool, it represents a starting point for discussion, framing what the tool is. This helped with the recruitment of companies and framing discussions around it.

Figure 7-2: The vis-UI-lise brand identity

The vis-UI-lise name and its written form were chosen to be easy to insert in text in a way that is self-descriptive and stands out. The logo builds on this with the use of a magnifying glass (See Figure 7-2 above). The gap at the top and bottom is an example of the Gestalt law of closure (Wagemans et al., 2012) hinting at the psychophysical nature of the approach.

7.2.2 Training Presentation

The training presentation consists of 58 slides (see Appendix B for further details) and a small selection from this is shown in Figure 7-3. The presentation is broken down into 4 sections as follows:

(1) What's the problem?

This section uses the iPhone example from Chapter 4 to highlight the visibility problem in the form of a common everyday device. Other examples are also given. This frames the next section about understanding vision.

(2) Understanding Vision

Vision is introduced through a series of examples of optical illusions and the ‘missing gorilla’ video is shown (See Section 1.1 in Chapter 1 for details). This is done to be both engaging and to challenge the problem of ‘naïve realism’ and frame vision as a psychophysical phenomenon. This progresses to the expression ‘believing is seeing’ as a succinct and memorable way of summarizing this (Green, 2008).

(3) Evaluating User Interface Visibility

The evaluation framework is introduced progressively using the car HVAC example (See Section 5.6 in Chapter 5) leading up to it being displayed in the form of the vis-UI-lise tool.

(4) Introduction to the tool

The vis-UI-lise tool is described with the associated support elements. These are discussed in the remainder of this section.

Figure 7-3: A selection of slides (6 from 58) from the training presentation

7.2.3 Evaluator Guide

The Evaluator Guide covers similar content to the presentation described previously. However, it contains detailed textual descriptions that ensure it stands alone and represents a reference document that can be consulted by practitioners who have seen the presentation. Sample pages from the 14-page document are shown below and details about accessing a full copy are available in Appendix C.

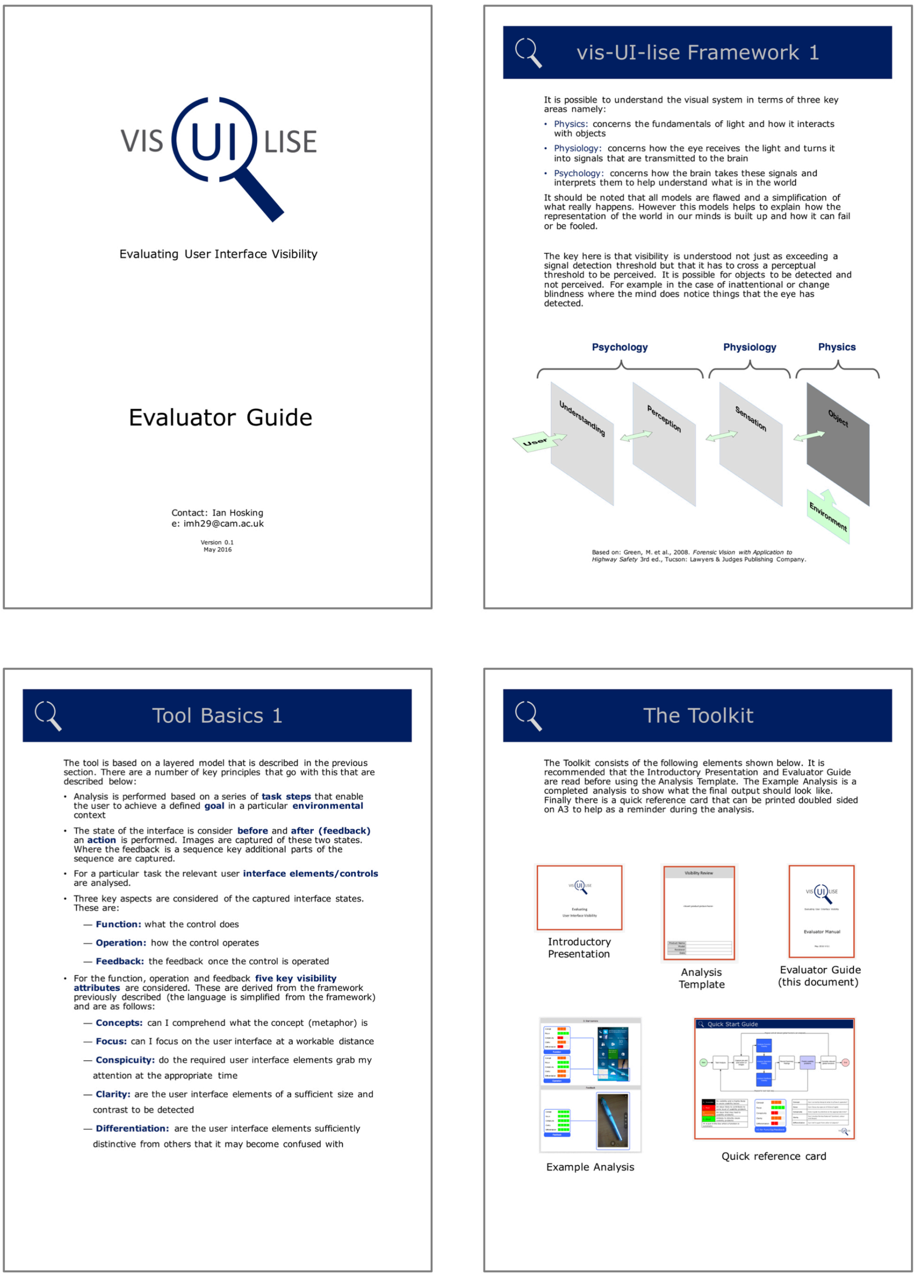


Figure 7-4: Sample pages (4 from 14) from the Evaluator Guide

7.2.4 Quick Start Guide

The Quick Start Guide is an A3 double-sided sheet shown in Figure 7-5 below and is available in full-size in Appendix D. The first side contains the evaluation process diagram (similar to Figure 5-15 in Chapter 5) combined with the 4-level rating scale from 'invisible' to 'good' and the 5 attributes that are rated for 'function', 'operation' and 'feedback' of an interface. The second side elaborates these stages of the interaction in the form that represent the key elements that have to be completed in the PowerPoint Evaluation template (see Appendix E). The quick start guide is aimed at being both an overview as well as a quick reference to completing the scoring.

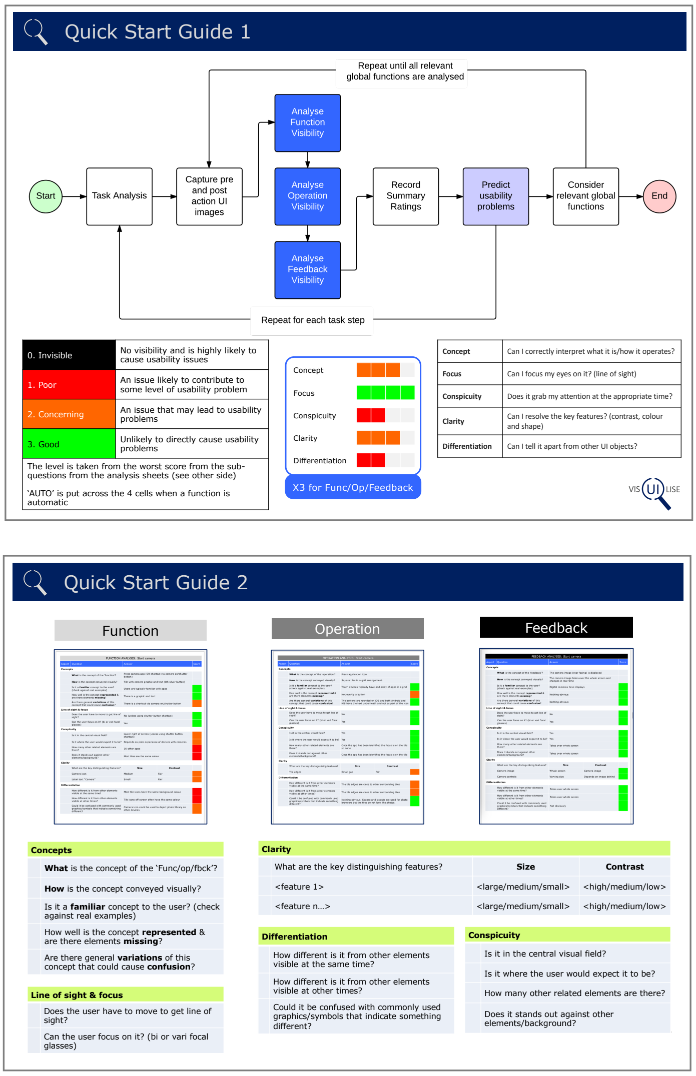


Figure 7-5: The double-sided A3 Quick Start Guide

7.2.5 vis-UI-lise Evaluation Template

The vis-UI-lise evaluation template is a 'blank' version of the slides (for further details see Appendix E). The aim is that the evaluator can start recording immediately into the template.

7.2.6 Practice Exercise

The practice exercise (for further details see Appendix F) is based on a subset of the output of the phone evaluation (Appendix A), reduced down to just steps 3 and 14. These were selected based on the range of evaluation issues they cover. The results are removed to enable people to practice completing an evaluation. The idea is that this can be used as part of practitioner training.

7.2.7 Phone Example

A full version of the phone analysis (Appendix A) is also given. This is helpful as a reference for the practice exercise but also as an example of a complete analysis so practitioners can see what a full evaluation looks like.

7.3 Method

7.3.1 Method Selection

To address the supplementary research question of 'how well and consistently will usability practitioners use such a tool to improve the visibility of user interfaces?' and the associated five requirement areas discussed earlier, requires an understanding the broader context of UEM use. This is particularly true of the issue of scope and positioning. This points to framing the evaluation as a socio-organizational concern (Dix, Finlay, Abowd and Beale, 2004), that cannot be reduced down to a simple experimental evaluation with a quantitative outcome. This then leads to considering sociological methods as a form of enquiry. As this is the first iteration, and limited work has been done at this stage on the broader evaluative context of practitioners, it points towards a more open enquiry afforded by a grounded approach (Creswell, 2009, pp.3-19). Grounded theory was defined by Glaser and Strauss (Glaser and Strauss, 1967) as the *"discovery of theory from data systematically obtained from social research"*. It contrasts with logically deduced theory generation from 'a priori' assumptions. It is an inductive approach discovering theory 'grounded' in the data.

Since this seminal work, grounded theory has diverged into differing approaches (Rieger, 2018). The work of Charmaz (Charmaz, 2011) acknowledges that the researcher does not engage as a 'tabula rasa' (blank slate) but they bring knowledge, experiences and interests that influence the framing of the enquiry. This leads to a 'constructivist' grounded theory progressing from the more objectivist, positivist origins. As such it assumes that the analytical outcomes provide *"interpretive renderings of a reality rather than an objective reporting of the reality"*. Charmaz's position is respectfully disputed by Glaser (Glaser, 2007) and it can also be argued that it conflicts with the author's critical realist ontological stance (discussed in Chapter 2) which does not extend to multiple realities. However, in practice, this constructivist approach is not inconsistent with the critical realist notion of multiple, interpretive perspectives on an ultimate reality. It resonates with the strong face validity that we are not a 'tabula rasa' and the pursuit of some faux objectivity seems unhelpful.

Regardless of these differences, Charmaz (Charmaz, 2011) provides a useful summary of the key aspects of all forms of grounded theory which are shown in Table 7-2 below.

Table 7-2: Key grounded theory principles from Charmaz (2011)

Principle
a) simultaneous data collection and analysis
b) pursuit of emergent themes through early data analysis
c) discovery of basic social processes within the data
d) inductive construction of abstract categories that explain and synthesize these processes
e) sampling to refine the categories through comparative processes
f) integration of categories into a theoretical framework that specifies causes, conditions, and consequences of the studied processes

Charmaz points out that the basic question driving all grounded theory, as set out by Glaser (Glaser, 1978) is, *"what is happening here?"* and this is what underpins this phase of research. The more detailed protocol for this research is split into two key stages. The first is the data collection and second is the analysis protocol. However, in practice these overlap and interact as in principle 'b' in Table 7-2 above, there was the *"pursuit of emergent themes"*, both within and across the interviews leading up the analysis of interview transcripts. It becomes *"ongoing involving continual reflection"* (Creswell, 2009, p.184) and is not merely reflective but ultimately reflexive in nature as it involves *"all aspects of ourselves and our contexts"* and therefore influences *"the way we research"* (Fook, 2015, pp.443-444).

7.3.2 Interview Protocol (Data Collection)

The interviews were semi-structured and followed Charmaz's (2011) description of them being *"a flexible, emergent technique; ideas and issues emerge during the interview, and the interviewer can then immediately pursue these leads".* There was a clear overall structure to the interviews that allowed *"issues"* to *"emerge"* and to *"immediately pursue"* them. This structure was in three broad stages of set-up, engagement, and preparation for further detailed analysis. These are outlined below:

Set-up

1. Recruitment

It was deemed key to get feedback from active practitioners who were strong candidates for using a tool like vis-UI-lise. In practice, such people are naturally very busy and therefore the research had to be presented in a way that was quick to understand and to see the potential value to them and their organisation. A short script or 'pitch' was written to ensure clear, concise and compelling communication with potential companies, whether by email or phone.

2. Confidentiality Agreements

Where there was the potential to use confidential projects as examples or to test vis-UI-lise on, confidentiality agreements were signed in advance of the research taking place.

3. 'vis-UI-lise Tool and Support' Delivery

The elements of the vis-UI-lise tool and the support were saved in a dedicated folder for each company in a cloud-based service (Dropbox™). This provided secure access during the research, which was given to the participants in advance of meeting with them.

Engagement

4. Briefing and Informed consent

As part of the recruitment, the participants were given a clear outline of how the research would work. This was repeated and elaborated at the start of the first meeting. This briefing included reviewing and signing an informed consent form which followed the same ethical approach as used in the user testing and described in section 6.5.2 of Chapter 6, which is consistent with the British Psychological Society's (2009) Code of Ethics and Conduct. After consent was gained the exact format of the engagement varied across the different companies but the following represent the core approach common to them all.

5. Training

Training consisted of using the support materials based around the introductory presentation described in Section 7.2.2. Effectively the presentation provided the structure for delivering the training and introduces the rest of the support materials.

6. Use

Once the company had been trained, the subsequent use of the vis-UI-lise tool was dependent on how it would fit in with their existing projects. This varied from immediate use, after the training, to later on when they could apply it to a 'live' project.

7. Follow-up

A follow-up call or meeting was arranged with three of the companies to see how they found using the vis-UI-lise tool. This was recorded for later transcription.

8. Note writing

During all verbal engagement with the companies, including phone calls, notes were taken.

Preparation for Analysis

9. Reflection

After all significant verbal engagements, time was taken to write up reflections on what had happened using the notes taken to aid the process. The aim was to do this as soon as possible afterwards, normally within a couple of hours after finishing, to maximise the opportunity to recall what had occurred. Such reflection would continue beyond this, in particular trying to put into succinct language what was occurring. These additional reflections would be recorded as they emerged.

10. Transcription

The recordings for interviews were sent for transcription by a professional transcription service.

The overall output consisted of interview notes, reflections and transcriptions. The specific outputs across all four companies are outlined in Section 7.4.

7.3.3 Analytical Protocol

Drawing on Creswell's (2009, p.185) summary of a more generic qualitative process and Charmaz's (2011) specific description of a constructivist grounded approach led to the following protocol for analysing the collected data:

1. Data Collation

All the data was collated into a set of electronic folders for each company. This included interview notes and transcripts from the engagements with the four companies.

2. Overview

All the data items where read sequentially to provide a sense of the whole from the first to the last engagement.

3. Coding

The interview transcripts were then coded according to Charmaz's (2011) approach that comprises of an initial open coding progressing to more selective coding as codes start to get reused and more dominant codes emerge. This was done manually using highlighting in the Microsoft Word™ documents and the codes were collated in a Microsoft Excel™ spreadsheet as the coding progressed. Using a dedicated analysis tool such as NVivo™ was considered but a manual approach was adopted as the number of interviews was relatively small. Dierckx de Casterle´ et al. (2011) describe an over-reliance on qualitative software as a major problem, as it can get in the way of the data, pushing users to early coding without reading and re-reading it first. Therefore, the simple approach taken here is appropriate.

4. Memo writing

As the coding progressed simple memos were written reflecting on what the codes where highlighting, and the relationships between codes and potential more abstracted categorisations for the codes. Memo writing is described by Charmaz (2011) as a *"crucial intermediate step that moves the analysis forward"* and this proved to be the case including helping refine the code hierarchy. In-line with Charmaz, memos were written quickly without aiming for 'perfection', but as part of the trajectory towards the themes.

5. Code rationalisation

Once the coding was finished, the code set was rationalised by two rounds of grouping using columns in the Microsoft Excel™ spreadsheet. This grouping was informed by the reflections from the memo writing.

6. Code framework

Using the grouping of the coding further refinement of the coding framework was performed by creating a hierarchical representation of the codes with an associated refinement of the language beyond the initial simple codes. This led to a set of top-level overarching concepts.

7. Themes & Model Development

All the proceeding steps lead to the development of key themes and the development of a model to link these themes together in a coherent way that addresses the purpose of the research, which is focussed on understanding practice and how vis-UI-lise fits within it.

8. Interpretation

Having finalised the analysis, the insights were interpreted to derive recommendations for improvement for the vis-UI-lise tool and represent the 'what next' in terms of further iterative development.

7.4 Data Collection

The following data was collected using the protocol outlined in section 7.3.2. vis-UI-lise was tested with four different companies. These are anonymised due to potential commercial sensitivity. A description of the different companies and an overview of what was done with them is as follow:

Company A

Overview: The company develops a diverse range of products across different markets, for different clients, with a strong emphasis on usability. This includes medical devices.

Approach: This was the first company and a pseudo-experimental approach was tried. Three practitioners were given the same phone and task sequence as used with the vis-UI-lise tool and asked to evaluate as they would for a client. The aim of this was to see what problems experts would find compared to vis-UI-lise. However, it became apparent that the amount of time this would take made it infeasible in a 'lab' style format and restricted the amount of time available to train participants on the vis-UI-lise tool. With the subsequent companies, the focussed shifted onto directly asking participants about their current practice and training on the tool for subsequent use on their projects.

Recording: Video recording with note-taking, followed by later reflection.

Company B

Overview: The company develops a range of healthcare products, for different clients, with a dedicated human-factors team.

Approach: The evaluation with Company B followed an approach that was modified from the lessons learnt from Company A. This consisted of distinct parts: (1) a phone call to set-up an initial training session that included some initial questions about their current practice; (2) a lunchtime presentation to eight staff for general feedback; followed by (3) a specific session regarding a particular project under a confidentiality agreement, with a sub-set of these staff. No recordings were made as it felt it was not appropriate with the confidential nature of the project. Interview notes were taken. (4) A follow-up interview with a key team member was then arranged to see how they got on applying it to the project in question.

Recording: Interview notes were taken for all engagements and an audio recording was made for the final interview which was then transcribed. Reflections were written up after each engagement.

Company C

Overview: The organisation develops products for people with disabilities.

Approach: A training session was held which included getting the participant 'started' on a current project, which was at a relatively early stage. Use was made of a functioning prototype device, which was well suited for analysis by vis-UI-lise and at a point that design changes could still be made. A follow-up video call was made to discuss how using the tool went.

Recording: Notes were made during the initial training session and the follow-up call was recorded and transcribed.

Company D

Overview: Company D was an independent human factors consultant working on a consultancy assignment, where their client agreed that the vis-UI-lise tool could be used.

Approach: The consultant was trained on the vis-UI-lise tool and then they used it on their client project. This was followed up with a face-to-face interview.

Recording: Notes were made during the initial training session and the follow-up meeting with the latter recorded for transcription.

The various outputs from the four companies are summarised in Table 7-3 as follows. This includes the number of practitioners directly engaged with at each company (shown as part of the header row).

Table 7-3: Data items generate for each of the four companies

Company A	3 Participants
Data 1	Video of workshop
Data 2	Spreadsheet output
Data 3	Post-workshop reflections
Data 4	Video transcripts

Company B	8 Participants
Data 5	Notes/reflections from set-up call
Data 6	Notes/reflections from the workshop
Data 7	Interview recording
Data 8	Transcription

Company C	2 Participants
Data 9	Recording of interview
Data 10	vis-UI-lise analysis PowerPoint
Data 11	Scan of written notes from post-interview reflections
Data 12	Interview reflections
Data 13	Transcription

Company D	1 Participant
Data 14	Training interview notes
Data 15	Interview recording
Data 16	Interview notes
Data 17	Interview reflections
Data 18	Transcription
Data 19	vis-UI-lise analysis PowerPoint

Individual practitioners are referred to in abbreviated form in terms of the company they were from and a number in sequence from the first to the last participant within that company, split by gender. So, the companies from A to D are referred to as CA, CB, CC, CD and the practitioners from F1 to F3 (Female one to three) and M1 to M3 (Male one to three). So, for example, CB-F2 refers to the second female practitioner from company B.

7.5 Analysis

The following analysis is based on the process described in Section 7.3.3 using the data gathered through the protocol outlined in 7.3.2. It should be stressed that although it is written in a linear form the process itself was very much driven by the *"pursuit of emergent themes"* (Charmaz, 2011) and *"continual reflection"* (Creswell, 2009, p.184), described earlier in Section 7.3.1, as part of the grounded approach.

7.5.1 Coding

The open coding process led to the following coding hierarchy shown in Figure 7-6. This was the output from the synthesis of the coding, memo writing and code rationalisation, that occurred throughout the analysis of transcripts and drawing from other data sources and experiences gained through engagement with the companies. The overarching theme is that of 'fit', in other words, how does any UEM tool fit within the process, culture and project requirements within a company. This is then split into codes relevant to any UEM tool which are grouped under the green headings. Issues related specifically to feedback on the vis-UI-lise tool are grouped under the single orange heading on the right side of the diagram.

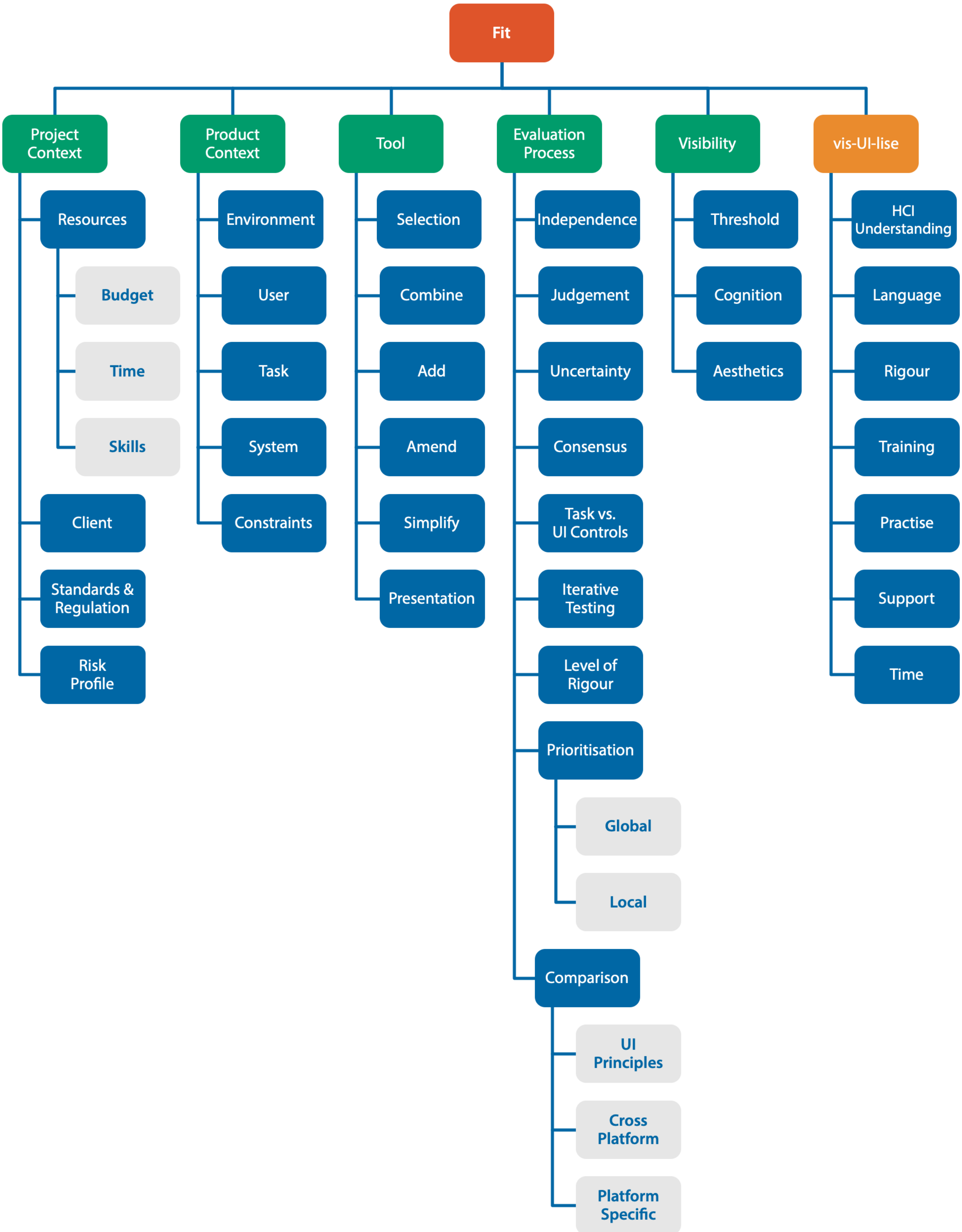


Figure 7-6: Hierarchical code framework

The question arises with regards to this form of analysis about using multiple coders to add diversity and insight into the coding process. Apart from the resource requirement of the time required, there is an additional barrier that the coding is not just about the specific interview transcripts but other contextual information concerning understanding the nature of the companies work and all other interactions leading up to the interviews such as the initial training with vis-UI-lise in three of the four companies. Therefore, it would be problematic for someone to simply code the interviews in isolation. Future work should, however, consider budgeting for additional resource to allow a diverse input. It is also worth noting that this is the first iteration so the level of analysis is sufficient to inform the next iteration which will be discussed in the latter part of this chapter.

7.5.2 Themes & Model Development

The analytical process leads to the development of key themes and a model to describe how they relate to each other. This again is an iterative process and the development of a model helps refine the themes. As was described earlier, this draws on the wider context to the interviews as well as the interviewer's experience of working in similar roles to those interviewed, and also as the developer of the vis-UI-lise tool. Such a lack of independence is likely to be an anathema to a positivist, but from the interpretivists' stand-point is an accepted reality of such an investigation and made transparent here accordingly. It should be stated that the underlying motivation is not to 'prove' that the tool is 'good' but to find areas for improvement in the next iterative cycle. This aim mitigates that lack of independence as finding problems is seen as a good thing.

In-line with the code framework, the themes are divided into those relating to practitioners' current practice of usability evaluation and those specifically related to issues with the vis-UI-lise tool (which are covered later in Section 7.5.4). There is some degree of overlap, but the split was deemed helpful in guiding the future development of the tool. The themes relating specifically to current practice are shown in Figure 7-7 under the headings of 'making the evaluation approach fit the context' and 'evaluation judgements'.

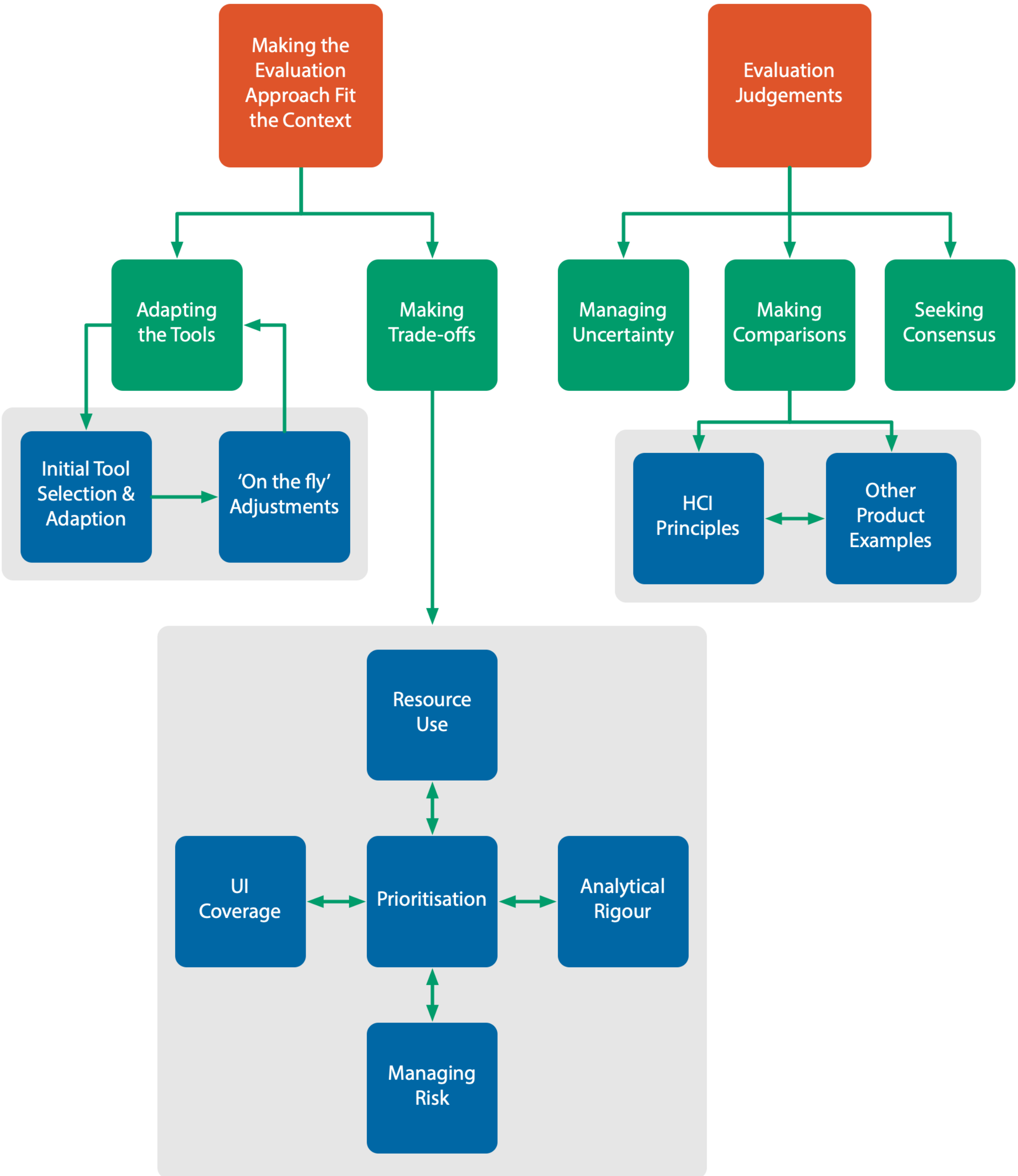


Figure 7-7: Themes hierarchy relating to the general process of usability evaluation

The model is broadly an 'explanatory' one in line with the types of models outlined by Rogers (2004) and discussed earlier in Chapter 3. A synthesis of the discussion around these themes at the green level in the hierarchy is described in the following section.

7.5.3 Current Practice

Making the Evaluation Approach Fit the Context

What becomes apparent fairly quickly is that much of the practitioners' conversation, regarding their approach to usability evaluation, is oriented around what the client wants or allows. As one practitioner put it "*It's kind of customised for each client*" (CA – F1). Indeed, a client can be integral to the evaluation process itself and as another practitioner put it when this happens *"the client is steering it"* (CB – F2).

What the client wants is not just about their personal views but the context of the 'product' and the 'project' which it sits within. There are a variety of contextual factors within these categories of product and project, which are in the codebook, shown previously in Figure 7-6, and elaborated and reordered further below.

- **Project Context**
  - Meeting the 'client' needs is not only a key driver but can also result in practitioners having a different view from that of the client about what is best for the product and its evaluation. *"There might be something that they're interested in that we don't think is an issue"* CA-F1.
  - Project 'resources' in terms of 'budget', 'time', and the available evaluation 'skills' of the team have a profound impact on the nature of the evaluation. *"It depends what you as the client wanted, whether you had the budget, and we'd look through it all or whether you just wanted to sit down for an hour and look at it."* CA-F1
  - The risk profile of a project and the associated product varies widely from, the strict formality of a User Failure Mode and Effects Analysis (uFMEA) on a medical device, to consumer products that are not

safety-critical. This contrast in the risk profile and its impact on the focus of the evaluation, was summarised clearly by one of the interviewees discussing the difference between medical and consumer products, *"consumer is the opposite; it's not there isn't necessarily use risk, so you're not picking tasks based on what's the worst that could go wrong, you're picking it based on commercial risk, what do you have to do to optimise the fact that people enjoy using it, but sometimes that's considered to be purely a marketing exercise, it doesn't actually go into interaction."*

- **Product Context**
  - The product context can be framed using the model discussed in Chapter 3, and represented in Figure 3-25, covering the 'user', the 'task' to be performed using the product, and the wider 'system' within which it sits. This model can be extended to include a range of specific constraints, which may include standards or regulations the product must comply with. This is particularly relevant for medical devices where regulation is tighter. Also, there may be technical constraints from the underlying software and hardware platforms used in the product development. Finally, the fidelity of the product available will vary from the purely conceptual to a finished product. With regard to the vis-UI-lise tool, one of the interviewees even inquired about the potential value in a pre-concept phase, *"Can you use it as a design tool as well as an assessment tool? That's my question because a lot of the stuff that we do is a blank page."* (CA-F1).

In summary, a UEM has to fit into a highly variable range of contexts consisting of a diverse range of contextual factors. The next two themes concern the strategies/approaches practitioners use to address this variability.

Adapting the Tools

Even when a company has a more rigorous process for medical device evaluation, there is still the need to adapt to the context, *"…so we have a template that we would be using for that, and we quite often adapt it per project because it's never quite right* (CB-F2). This adaption is not necessarily confined to single tools but a project may be making use of different tools and combine their outputs, *"and in terms of other tools there may be some things that we may have done separately, and there was on this project in terms of steady-state analysis or various other things that we might be drawing from."* (CB-F2).

In addition to this 'macro' level adaption, made during the selection and upfront adaption of the tools, there is also 'micro' adaption as the evaluation unfolds. To an extent, this happens 'on the fly' and is particularly true with regards to focussing in on key areas of concerns, or *"showstoppers"* (CC-F3), as one practitioner described them. Also, as a design evolves it may be retested and there is a need to capture this, not only to create an audit trail, but also to use the previous analysis to inform the current one, *"you add the extra columns* (spreadsheet record) *for your study every time you do it, which is saying you've got your previous, what the potential mitigations would be from the previous study that you're trying to resolve, and then within that, you're saying 'Okay, well within this study what issues have I continued to see? What new issues have I got, and which ones have been mitigated by the design change that you've made?' and you just try and capture that."* (CB-F2). This shifts the emphasis from the evaluation producing a static output, at a fixed point in time, to one that is a *"live document"* (CB-F2), which not only helps the on-going process but also forms a *"narrative"* (CB-F2) of what happened as a summative report at the end of the project.

Making Trade-offs

In the same way that client management is a thread running through a project, so is making trade-offs. Indeed, the two are closely related as the client is a key factor in how trade-offs are made. The following quote relates to a specific discussion around the tension between industrial designers and usability practitioners, but also typifies the general issue of the need for trade-offs, *"…and I've seen it everywhere and I don't think <Company B name> is different in any way, shape or form, and it depends on the specific designers you're working with, some are more open to usability being a more important driver than fitting in with the design style guide or whatever it is. And others don't see the issues in the same way for usability and think that you just can't do what we're trying to suggest or it's not as relevant or it's not as important, so they just don't put the weight behind it and so that takes them down a slightly different path of what the solution or the result should be."*

Figures 7-6 and 7-7, shown earlier, have the key parameters in the wider trade-off space and the setting of priorities around these. Resource use (budget, time & skills) is a key limiting factor that other criteria must be 'juggled' against. A key way of reducing resource use it to reduce the amount of the user interface that is covered and the level of rigour that is applied to each part of the UI. This relates closely to the 'macro' and 'micro' level adaption discussed earlier. Although such a reduction in scope reduces resource use, it then increases risk due to things that might be missed. The approach to risk was described earlier, with the contrast between the safety-critical nature of medical devices versus the lesser demands of consumer products. The latter can lead to a much more superficial evaluation being undertaken. Even with medical devices, the depth and breadth of coverage is still likely to be constrained by the available resources.

Evaluation Judgements

Making judgements occurs across all areas of a project and is part of the process of making trade-offs discussed above. It also occurs specifically within usability evaluation. The reliance on judgment over detailed methods was something that 'stood out' to the researcher during the sessions themselves. This included direct observation of the three designers from Company A performing an evaluation. With further reflection, as part of the subsequent analysis, it is not surprising judgements are used to enable efficiency within the constraints of the project, in particular time and budget.

Evaluation judgement is broken down into three themes (in green in Figure 7-7), which were derived from the interviews. The first concerns a driver which is to do with **'managing uncertainty'**. This relates specifically to uncertainty around what the user interface will finally be, the nature of the end-users and how they may use it. This uncertainty will typically be much higher earlier in the design process as the product concept evolves. The iterative nature of the testing, going as far back as the earliest development stages, has already been described above, and making a 'judgement' is a pragmatic way of determining what the issues might be, by 'imagining' what the interface could be like or behave.

The second relates to how such judgements can be informed by **'making comparisons'** with relevant, existing interfaces in different products. These may be relevant, for example, where a comparison is made with a similar product that addresses identical or similar user requirements. Alternatively, it may be compared with a particular user interface component that delivers the required function, for example, a mechanism for entering a date. Where a project is using a particular operating system and UI development environment, then a comparison can be made to other products that use the same user interface framework.

The third concerns the qualitative nature of a judgement that can lead to disagreement and the need to **'seek consensus'**. This could be described as an 'uncertainty spectrum', which ideally requires resolution through reaching consensus. The following quote again relates to the tension between the industrial and user interface design but articulates this wider issue.

> "*It's never normally as black and white as he's wrong and I'm right, it's just a greyness in between, so it's just almost like a dial of how far you are sort of going to push it towards usability versus the design intent.*" (CB-F2).

Current Practice Summary

The picture that emerges from current practice, is one where a UEM tool needs to be flexible to fit in with project contexts, which can vary significantly. Not only that, even within a particular project context the needs change through the project, from early low-fidelity concepts to the final design. Practitioners have to adapt tools to meet the needs of a project. They appear to have to make regular trade-offs and rely heavily on making judgements based on experience. This dynamic, fluid situation occurs at a 'macro' and 'micro' level within a project, from making structural decisions about how to shape an evaluation, to 'on the fly' decisions as the evaluation unfolds.

7.5.4 vis-UI-lise Issues

A set of themes were developed specifically to address issues practitioners faced with vis-UI-lise, after using it. The themes are broken down into two broad areas, 'understanding the tool' and 'operating the tool'. These themes were informed through general reflection across the various points of interactions with practitioners, as well as specifically through semi-structured questioning focussing on the language used in the tool itself. This questioning came about, as it became apparent that the descriptive language used in the vis-UI-lise tool was a key problem for practitioners. These themes and the relationship to each other are shown in Figure 7-8 below.

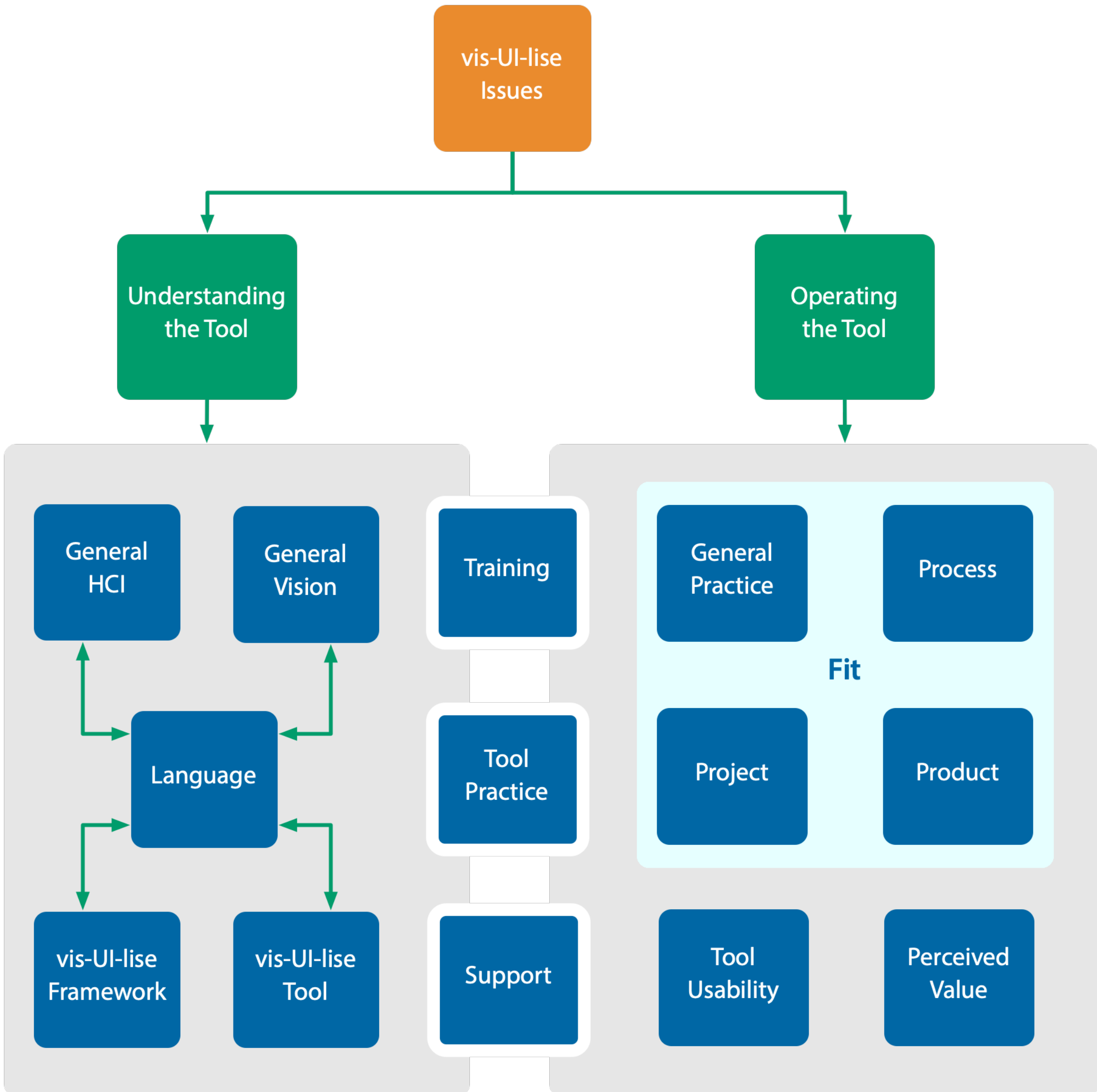


Figure 7-8: Themes hierarchy relating to the vis-UI-lise tool

Understanding the Tool

The issue of language, particularly concerning the hurdle questions (see Table 5-5), dominated the understanding of the vis-UI-lise tool. The language for the **vis-UI-lise framework** and **tool** was derived from the respective domains of **human-computer interaction** and **vision science**. This is shown on the left-hand side of Figure 7-8. Usability practitioners' difficulties were evident when they said they, *"struggled…to get to grips with these terms"* (CC-F3), meaning the hurdle questions and *"I was working very hard cognitively to make my own interpretation of what the questions were asking"* (CD-M3).

This feedback led to the questioning homing in on specific difficulties with the key terms framing the hurdle questions namely: 'concepts', 'line of sight and focus', 'clarity', 'differentiation' and 'conspicuity'. The second page of the Quick Start Guide (see Appendix D) was used to aid the discussion with the practitioners. A series of key responses from practitioners were tabulated to help inform potential improvements in both language and the presentation of the underlying meaning. Details about accessing this table are available in Appendix G, and an analysis of them is provided as follows.

The word 'concept' caused the most difficulties and is also the most important as it frames the evaluation of each task step. This is something recognised by one of the practitioners:

> *"if you get the concept stuff right, given that it's the first thing that the user - user as in me - will come to, I think if you get that right, it gives people confidence that they're going to understand the rest of the questions."*

The potential of other words was explored. However, the overarching problem is how to get to a description that is essentially independent of a control or function's current visual representation, in order to be able to assess whether this visual representation is effective. In other words, what is the word or phrase that describes the control or function independently of how it looks? The word 'function', for example, clashes with the need to distinguish between the underlying function, how that function operates and what the feedback is when it has been operated. One of the practitioners suggested splitting the descriptions out into three:

> *"I kind of wonder whether it's specific to those three types. So, you analyse function visibility, analyse operations visibility, analyse feedback visibility and I'm not sure that the word concept works across all three."*

This seems a sensible approach, and leads to the suggested potential wording for the three questions given below:

1. What does the underlying function/control do? (Function)
2. How is this function/control operated? (Operation)
3. How is the successful operation of the function/control communicated to the user? (Feedback)

Future work will be discussed in the next chapter, but this revised thinking could form the basis of a co-creation exercise with practitioners to devise more effective wording.

Issues with the word 'focus' concerned the potential confusion between focus as *"attention"* (CB-F2) and focussing with your eyes. However, another practitioner said that in the context of the hurdle question it is clear. Similar problems occur with the word 'clarity' i.e. between it being in the *"cognitive"* sense (CB-F2) or in terms of vision. The suggestion of calling it *"visual clarity"* (CC-F3) makes a lot of sense and again something that can be an input to future work.

The term 'differentiation' caused potential confusion with how it is different from 'clarity'. This is a problem of 'clarity' being defined as the basic ability for the visual systems to resolve the image in isolation, as opposed to how the image is differentiated from those visible at the same or other times.

Using ‘differentiation’ in the context of comparing it to other controls present at the same time is potentially much easier, as seen in the example below:

> *“I can immediately picture that if I was looking at a CD player and you said, ‘Differentiation. How different is it from other elements visible at the same time?’. ‘Well, it’s rubbish. There are 15 silver round buttons, and they’re all in a row’. I can immediately get the sense of that and understand how to answer the question.”* (CC-F3)

This practitioner did go onto comment that it is less clear where items are on other screens, i.e. not visible at the same time. This is something that could be further clarified through examples.

Finally, the term ‘conspicuity’ highlights the overall problem here of ‘technical precision’ versus ‘accessibility of language’. One of the practitioners summed this dilemma up well:

> *“It’s not a lovely word, but it’s probably the right one. …It depends if you’re willing to substitute it for something that might be less precise but might get you a more precise answer.”* (CC-F3)

An alternative description is a less elegant, but potentially more accessible expression, of *“attention-grabbing”* (CB-F2). So, ‘conspicuity’ becomes ‘attention grabbiness’, and is potentially much more understandable, albeit clumsy sounding.

The complexity of language and the dilemmas around choices, led to two of the practitioners going on to comment, *“…which, of course, would get easier with practice”* (CD-M3) and *“…so I wonder if it’s simply a case of more practice”* (CC-F3). This raises the question of whether they would persevere, particularly in the context of a pressured project, where they have established methods which they are comfortable with. This issue will be returned to when considering the theme of **'perceived value'**.

Related to the struggles with language is the associated support that the vis-UI-lise tool comes with. One of the participants commented on the value of the supporting documents and the 'proximity' of this at the point of evaluation:

> *"I think where you've got the visibility rating scale, for example, where you say all of the blue speech bubbles that say things like, 'Can I focus my eyes on it?' which describe the term 'focus'. 'Does it grab my attention when required?' which describes the term 'conspicuity'. That's really helpful because you have to hand at any given moment all of the definitions of the terms. Maybe doing more of that throughout where you can. But I think you often are doing that. But certainly, where that's done, it is helpful, where you've got the explanations to hand the whole time."* (CC-F3)

In further discussions, about the use of examples to explain the hurdle questions, they confirmed the potential value of this and having them to hand at the point of evaluation: *"Yes. They go a really long way because it's so immediate."* (CC-F3)

The is clear potential to improve the support and to look at ways to encourage practitioners to persevere and develop mastery of the tool. Such changes would need to be integrated into the training and the three themes of **'training'**, **'support'** and **'tool practice'** can be seen as a bridge between **'understanding the tool'** and **'operating the tool'** as shown in the middle of Figure 7-8.

Tool Operation

The ability to make any tool fit a specific context of a project is likely to be critical in its success. **'Fit'** is described in terms of four related themes which are contained within the white box, on the right side, of Figure 7-8. Two of these **'product'** and **'project'** help described and inform the constraints and indeed opportunities that are typically brought by a client or product owner. The second two of **'process'** and **'general practice'** relate to the team delivering the **'project'** and associated **'product'**.

The analysis so far has covered a range of issues around **'project'**, **'product'** and the evaluation **'process'**. However, further analysis of **'general practice'** reveals some interesting dynamics. We have already considered **'seeking consensus'** and what can underly the differences in approach across practitioners. This then relates to the overall corporate practice that may be embodied in formal processes or informal ways of doing things. The complexity of this in terms of a social system can be described as a 'community of practice' (Wenger, 1999). A greater difference is likely to occur across disciplines, such as those described earlier for industrial design and usability. These may be narrower within a discipline like the one described below.

> *"The people that I work within the HF* [Human Factors] *group we are all of a quite similar mindset and we are very open to each other's views"* (CB-F2)

Regardless of the level of the differences, there is a *"mindset"* both 'corporate' and 'personal' that is important to understand to determine the most appropriate format for a UEM, how to 'sell' it and the training for it.

Although **'fit'** is key and challenging, it also represents an opportunity. One such opportunity, that was highlighted, is fitting vis-UI-lise in with the standard PCA (Perception-Cognition-Action) framework used in medical device evaluation:

> *"…it felt quite familiar with what I was trying to achieve because it fits in with the whole aspect of the PCA and the thinking and breaking down a task to try and sort of assess it."* (CB-F2)

The fit here is potentially strong for two reasons. Firstly, the perception-cognition-action sequence is aligned with the cognitive approach used in vis-UI-lise. Indeed, there is the potential to help with the ambiguous boundaries in the PCA approach, between perception and cognition, which were highlighted by a couple of practitioners:

> *"I often go 'Is that perception or is that cognition?' and sometimes it's both and I'll just put it under perception or cognition, randomly."* (CA-F1)

> *"I think it would be good to use this…where a lot of it is visual, or the overlap between cognition and visual, I think it works really well for…it would help me be a bit more specific."* (CB-F2)

As the latter quote shows, there is an opportunity for vis-UI-lise to bring greater insight and clarity to a PCA analysis. In addition, such an analysis typically requires a task analysis to be performed, which is also the case when using vis-UI-lise. This level of process overlap means that vis-UI-lise could potentially be integrated as part of a PCA analysis.

Even if a good overall fit can be established, the basic usability of the vis-UI-lise tool is important to ensure a high level of productivity. This proved an issue in its current form that uses PowerPoint. This was chosen because it is widely available and easy to develop the first version of the tool in. However, it did cause problems:

> *"My experience of using it is somewhat wrapped up in the input mechanism… using PowerPoint to enter the data, which I had some usability issues with"* (CD-M3)

In this particular case, the practitioner did relate it to his specific computer set-up, but none the less it is a more general area of concern and needs to be addressed.

So far, the orientation of the analysis is to highlight and understand issues, to drive improvement of the tool. It is also useful to consider the positive feedback and how the 'perception of value' may help in encouraging uptake. The following discussion typifies the views expressed around the rigour and detailed insights the vis-UI-lise tool can bring:

> *"It depends what you're trying to do. If you want a diagnosis of where the problems are, it's very helpful, if you just want an overview of whether it's any good or not it's less helpful. But your tool was definitely at the diagnostic end of the spectrum, which is detailed, rigorous, and yeah, powerful, but not quick and easy to apply"* (CD-M3)

> *"So what do you mean by the word powerful? What's powerful about it?"* (Interviewer)

> *"I think it forces the assessor to think about things in a way that I certainly wouldn't have done before, so it asks questions of you that you would not have thought about. And those questions can be revealing of problems."* (CD-M3)

Another practitioner summarised it as follows:

> *"where you talk about it being a systematic tool, it absolutely is. …the thoroughness and the systematic approach is really constructive."* (CC-F3)

The first series of quotes above, also highlights once more the issue of 'fit', with the expressions, *"it depends what you're trying do"* and the *"spectrum"* of need.

Finally, there was a specific instance of 'actual' value where the use of vis-UI-lise highlighted a problem during the development phase, which proved to be a significant problem when the product was tested later with real users.

> *"People did have exactly the problem that was highlighted using your tool, which was the confusion between a short press of the power button, which switched it between asleep and awake, and a long press of the power button, which switched it between powered off and powered on. Because that length of time isn't defined, and because people didn't appreciate the two functions, as you absolutely flagged up that would happen, there was confusion. This tool did precisely predict what people were going to do in that instance. Interestingly, it wasn't primarily about the visibility. It was about not being able to understand what the function did because it wasn't well described, I think. But the tool picked it up, absolutely."* (CC-F3)

There are two interesting points with this comment. Firstly, at the time of evaluation, the practitioner struggled to unpack the specific issues related to the power button, in particular, its support of multiple functions. In took input from the researcher at the time to draw this out. The question that arises is how much was this to do with the tool, and the language issues discussed previously, and how much due to the inherent complexity of the product functionality involved? This is something that needs further work to unpack. It is interesting with insights gained from user testing that the practitioner was then able to articulate what was going on in a way that was not evident at the earlier evaluation stage.

Secondly, and related to the first point is that they say that *"it wasn't primarily about visibility",* this is counter to the whole notion of it being a psychophysical phenomenon and implies that it is hard for practitioners to relate to. This struggle also relates to the issues concerning the ambiguous boundary between cognition and perception discussed earlier.

The last example highlights an underlying issue around the level of understanding practitioners have of human-computer interaction. This is something that the researcher felt during the delivery of the training and subsequent issues that arose. A certain level of knowledge was assumed during the development of the vis-UI-lise tool and this may have been over-optimistic. Therefore, further work is required to determine the variation in HCI knowledge across practitioners to help determine how best to improve the tool, its support and training.

vis-UI-lise Issues Summary

The vis-UI-lise tool offers a high level of rigour compared to the practitioner's current approaches. This is even true for the two companies that develop and evaluate medical devices under a tight regulatory framework. Although this rigour was valued, practitioners are probably sufficiently happy with their current approaches to consider changing them. Therefore, there needs to be a compelling reason for them to change. Even in the case where the tool predicted a significant problem, well in advance of user testing, it is not clear if the practitioner would adopt the tool if it was made available. However, the tool does show significant promise and there are a number of things that can be addressed to help improve the usability of vis-UI-lise and deliver clearer, tangible value. This is likely to help with adoption. These are discussed in the next section.

7.6 Implications for vis-UI-lise

Reflection on the work with the four companies leads to the following proposal regarding high-level objectives for the next iteration within the DRM framework. They are as follows in Table 7-4.

Table 7-4: Proposed objectives for future development mapped to the high-level requirements for a UEM

	Objective	Mapping to Requirements
1	**Tool Usability:** Improve the overall usability of the tool to improve its efficiency	Usability
2	**Tool Adaptability:** Make it suitable for rapid adaption to different project contexts	Scope & Positioning
3	**Tool Variants:** Produce variants suitable for different domains, such as medical device development	Scope & Positioning
4	**Tool Development:** Make practitioners an integral part of the redesign and development process	All
5	**Tool Adoption:** Demonstrate the 'downstream value' to encourage adoption e.g. its ability to spot problems early	Persuasive Power

These objectives map to the high-level requirements outlined in section 7-2. Specific insights that inform these objectives are as follows:

7.6.1 Tool Usability (Objective 1)

At the core of the vis-UI-lise tool are the hurdle questions that present tool usability issues, due to their inherent complexity and wording. Drawing from the discussion in the earlier section on 'tool operation', Table 7-5 below highlights the problematic terms and suggested alternative wording. These represent a potential starting point for co-design with practitioners in-line with Objective 4.

Table 7-5: Problematic hurdle question terms, with suggested alternatives, based on interviews with practitioners

Current Term	Suggestions
Concept	Have individual definitions for 'function', 'operation', and 'feedback' as follows: 1. What does the underlying function/control do? (Function) 2. How is this function/control operated? (Operation) 3. How is the successful operation of the function/control communicated to the user? (Feedback)
Clarity	To avoid potential confusion with 'cognitive' clarity call it 'visual clarity'
Conspicuity	Augment or replace this term with 'attention-grabbing level' or 'noticeable-ness'
Differentiation	Consider using 'distinctiveness' instead

Two further ideas, inspired by the practitioners, are first, to have examples integrated with the support material e.g. for each of the hurdle questions. This could build on the car HVAC example shown in Figure 5-6 of Chapter 5. This in conjunction with the changes in language could make a substantive difference.

Second, provide a simple graphical representation of a cross-section of a human head, with the brain and eye emphasised interacting with a user interface, to show the psychophysical nature of vision. The diagram could then have the hurdle questions mapped on according to the staged nature of vision. This would build off the diagrams shown in Figures 3-23 and 5-5, in Chapters 3 and 5 respectively.

7.6.2 Tool Adaptability (Objective 2)

The user of PowerPoint offers a high degree of customisation but having specific examples and templates would make it easier for the practitioners to both see potential adaptions and also aid them in doing it. It should be noted that PowerPoint proved to be relatively weak in terms of the usability of data entry. This potentially could be fixed using macros to help automate this.

Regardless of the specific underlying tool. Rapid adaptions in the following areas need to be addressed:

- The ability to vary the coverage and depth of task steps e.g. miss steps out and vary the level of depth of assessment on a per-step basis with a clear recording of why this has been done.
- Have the option to record the evaluation of multiple versions of the product (design iterations).
- Ensure the document can be used in a 'live' manner by multiple practitioners with an audit trail of who has added and amended any of the content. This could take advantage of cloud-based document sharing, which can support shared viewing and editing.
- Consider ways the tool can help build consensus through the synthesis of different views.

7.6.3 Tool Variants (Objective 3)

Making the tool easy to adapt goes a long way to addressing the variability of need across different product and project contexts. However, this can be augmented by specific variants of the tool that could target the particularities of different domains. Three initial suggestions for this are as follows:

1. Medical Devices

Medical device development and usability evaluation centres around the perception-cognition-action model (CENELEC, 2015) and the use of uFMEAs, which stands for use Failure Modes and Effect Analysis (CENELEC, 2016). The practitioners from the two companies, that worked on medical devices, could see how this could be integrated with their existing practice. Working on such an integration is an obvious next step and would aid adoption, which will be discussed in the next section. Integration with a PCA/uFMEA addresses not only the medical device domain but more rigorous assessments in general, particularly in safety-critical applications.

2. Usability Heuristics

As was mentioned by the practitioners in Company A, a consumer product can have a very different emphasis and scope in terms of the breadth and depth of evaluation, compared to something like a medical device. Also, the use of evaluation heuristics seemed to be a 'go-to' aid for an evaluation. Therefore, augmenting something like Nielsen and Mack's (1994) heuristics is an interesting option. The importance of visibility in these heuristics was previously explored in Chapter 3 and to a degree, the hurdle questions apply broadly across all of the heuristics, and overtly with regards to something like the 'visibility of system status'. However, a more detailed investigation shows how four of the five hurdle questions map to specific heuristics. Figure 7-9 shows this mapping, using a keyword from each of the 5 primary hurdle questions (See Table 5-5 for the full questions). Note that 'focus' does not map directly to a specific heuristic but is important to a number of them. The heuristics are reordered to highlight the questions that map to the hurdle questions.

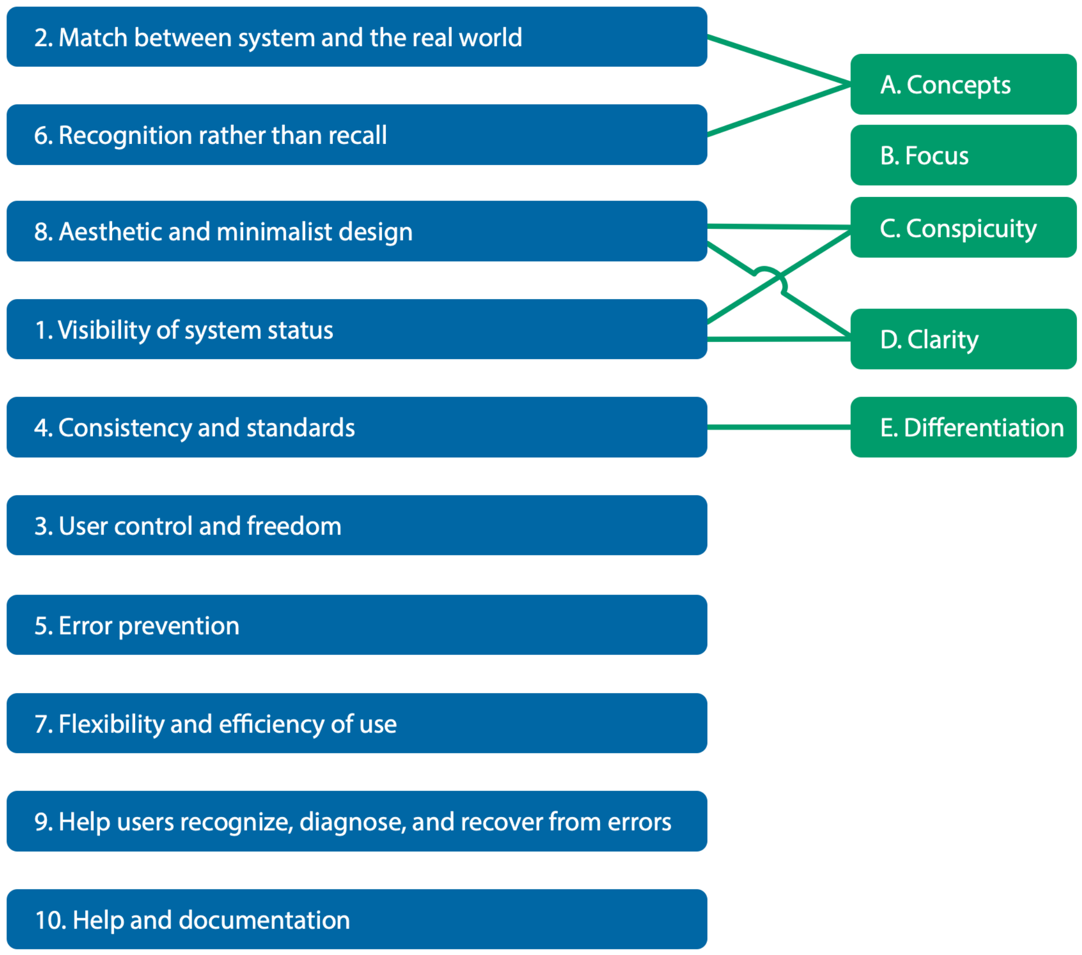


Figure 7-9: Primary mapping of Nielsen & Mack's (1994) heuristics to the hurdle questions

Of particular note is the link between the second heuristic, 'match between system and real world'. The associated description for this, provided by Nielsen and Mack (1994), encompasses much of the essence of the 'concepts' hurdle:

> *The system should speak the users' language, with words, phrases and* **concepts** *familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.*

More work is required to create an effective integration that may be a mixture of modifications and additions, but the potential is to create something that is an easy, incremental change for practitioners. As such, this mapping could form the basis of developing a revised or augmented set of heuristics.

3. User Interface Visibility Awareness

One of the insights from the interviews and observations was the reliance on judgement. Something that could build on this and be complementary to the other two options described above is to create a greater awareness of the cognitive nature of UI visibility. Indeed, this is already part of the training, but further work is required to understand what 'understanding' this generates within individual practitioners. One possibility is to create a simple 'vision spotlight' simulator (See Figure 3-17 from Chapter 3). This could be in the form of simple tubes of a specific diameter and length that restrict a practitioner's vision. The tubes could be made into a simple binocular style format or integrated with safety glasses. This would potentially do two things. Firstly, challenge the naïve model of vision and secondly, to simulate the searchlight nature of vision.

7.6.4 Tool Development (Objective 4)

So far, a range of requirements and potential changes have been described covering usability, adaptability, and use specific variants. The natural temptation is to just make these changes. However, it is a good point to reflect on what the most appropriate process for doing this would be. Two things standout in this regard. Firstly, having the vis-UI-lise tool has provided a powerful lens, in conjunction with practitioners, on the needs of practitioners. This observation is consistent with the view that design involves a co-evolution between the problem and solution space (Lou Maher, Poon and Boulanger, 1996; Dorst and Cross, 2001). Secondly, the needs of practitioners are highly context-specific and nuanced. Both of these issues could be well addressed by adopting a co-creation/co-production model (Chathoth et al., 2013; Brocklehurst et al., 2019; Voorberg, Bekkers and Tummers, 2014) for the next design iteration.

7.6.5 Tool Adoption (Objective 5)

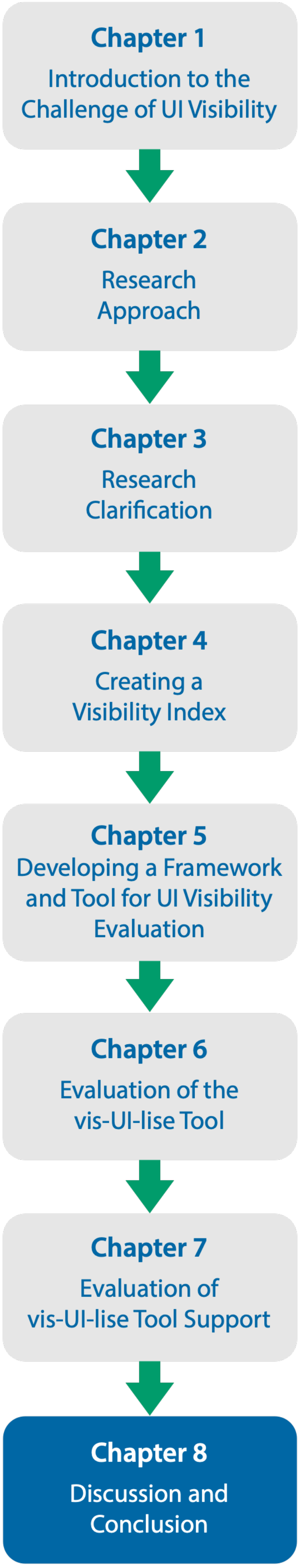
The implication from the work with the four companies is that practitioners may prove resistant to changing their existing practices that suffice for their current needs. The changes proposed above should all help in adoption. In particular, the tighter integration with existing approaches such as PCA/uFMEA and Nielsen and Mack's (1994) heuristics would hopefully 'shorten the bridge' between current and a modified approach using vis-UI-lise. The key is to show a business benefit that will centre on time and cost savings related to finding problems earlier and more quickly or avoid missing problems that could prove costly. Building evidence for this will be important to convince practitioners. Finally, if the co-creation/co-production approach was taken, it would create a set of early adopters and associated case-studies that could be used to promote the vis-UI-lise approach, in whatever form it may subsequently take.

7.7 Summary

The Descriptive II stage of the DRM focussed on addressing research question SRQ 3c:

> *Can and will practitioners use such a tool to reliably improve the visibility of user interfaces?*

Through working with four different companies, significant insights were generated with regards to general issues around usability evaluation and more specifically with the vis-UI-lise tool. The research helped to unpack the complex and varied world of usability evaluation. It also highlighted a specific instance of where the vis-UI-lise tool predicted a significant problem in advance of user testing. In addition, practitioners commented on the rigour that vis-UI-lise brings and leads to insights that they may have overlooked with their normal practice. This builds on the promise shown in the previous experimental study. However, the 'variability' of the needs uncovered across different product and project contexts needs to be met with 'adaptability'. To do this, and realise vis-UI-lise's potential, a number of significant changes have been proposed to address the issues found across the companies. So, as it stands with regard to the question, there is sufficient evidence to believe, with further work, that a variant of vis-UI-lise could augment existing practice to produce more visible interfaces.


Chapter 1
Introduction to the Challenge of UI Visibility
Chapter 2
Research Approach
Chapter 3
Research Clarification
Chapter 4
Creating a Visibility Index
Chapter 5
Developing a Framework and Tool for UI Visibility Evaluation
Chapter 6
Evaluation of the vis-UI-lise Tool
Chapter 7
Evaluation of vis-UI-lise Tool Support
Chapter 8
Discussion and Conclusion

CHAPTER 8

Discussion and Conclusion

8.1 Discussion of the Research

The early stages of this research considered the visibility of an iPhone 7™ running iOS 10™. The 'visibility index' for this, which just considered the navigation around the various home screens, showed that only 8% of the 622 functions available were visible at the top level. Even when repeated functions are excluded this only rises to 17%. Bearing in mind that this is only the functionality up to the point of launching an application, and not any functionality within them, it paints a stark picture regarding the lack of visibility. Moving forward a few years to an iPhone X™ running iOS13.2™ things have become even worse regarding visibility. The home button has now been removed (compare Figure 4-4 to 8-1 below).

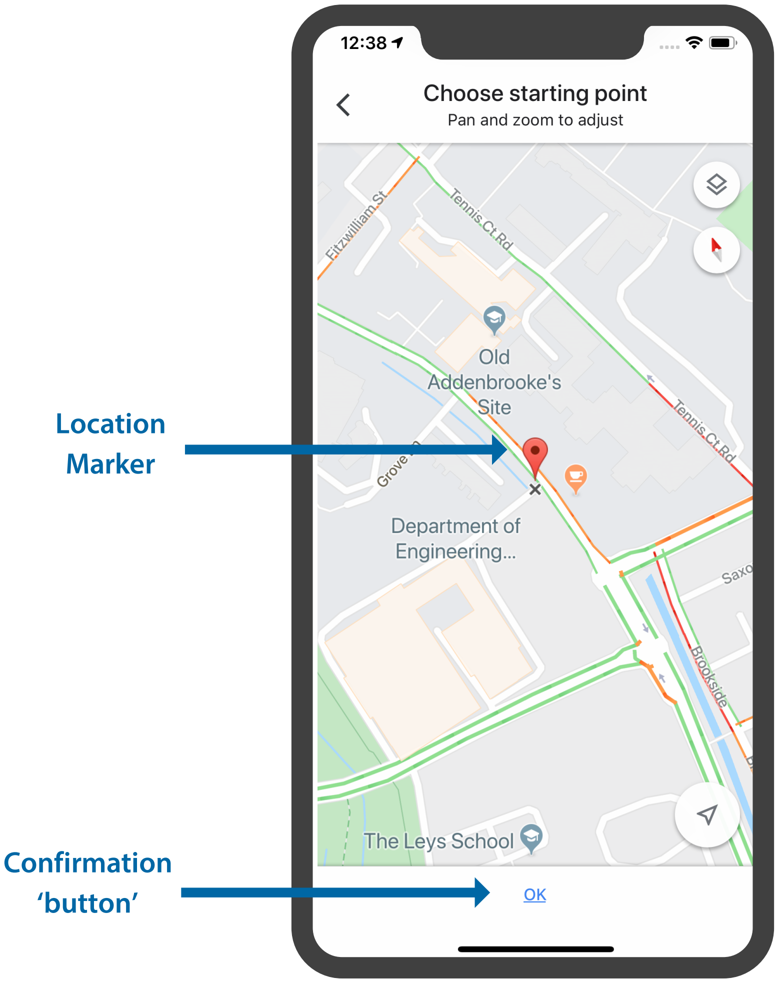


Figure 8-1: A screenshot showing setting a starting point location marker for getting directions on Google Maps (phone body is drawn graphically to represent phone hardware)

From a UI perspective, 'the button' it is both physically 'missing' and the alternative mechanisms are 'missing' too from a visibility perspective as it has been replaced by a combination of Face ID, to unlock the phone, and on-screen gestures to replace the other functionality the button provided (Apple Inc., 2019, pp.36-37). Figure 8-1 shows a screenshot for Google Maps™ (Google Inc., 2019b). The screen shows the process of manually setting a location to get directions. Two things stand out about this, or more explicitly do not stand out. Firstly, confirming that the marker (highlighted with the arrow in the middle of the screen) is in the desired place, is performed using the 'OK' button/link at the bottom of the screen (highlighted with the lower arrow). Figure 8-1 displayed in a document like this does not fully convey just how hard it is to see the 'OK' button in practice when the phone is held in the hand. The 'button' is small, not in the attentional visual spotlight and could also be obscured by the user's hand. It can easily go 'missing'. Secondly, the behaviour of the marker it unusual as the user has to move the map around while the marker remains centred. This is in contrast to the web version (Google Inc., 2019a) where the marker is dragged around the map. Here the functionality could easily be 'misunderstood', albeit it can be argued that is it relatively easy to work out what is happening as the feedback is instantaneous.

With this up to date phone example, we see again the 'missing', 'missed', and 'misunderstood' aspects of UI visibility. These aspects were brought together through the Three M's Model that describes a spectrum from the invisible to the visible (See Figure 3-10). This frames the problem and was fundamental to the research in moving the perspective away from one of a simple visibility threshold—a can I or can I not see it issue—to a multifaceted phenomenon.

This simple model represents an answer to the first of the ten high-level, guiding questions, namely, ‘what is the problem?’. The ten questions from Chapter 1 are shown below and will form the structure for the remainder of this section.

1. What is the problem?
2. Is it worth solving?
3. Can it be suitably framed as a question suitable for academic enquiry?
4. How well does the literature already answer the question?
5. What is the gap?
6. What conceptual understanding is required to study the gap?
7. What is an appropriate empirical approach (study) to address it?
8. What do the results tell us?
9. What is the contribution to the literature?
10. What else needs to be done?

Having established and framed the problem it is important to understand the value of solving it. One only has to close your eyes and use a typical digital device to highlight how critical visibility it. The example given in Figure 8-1 is typical of all the examples used throughout the work, in that it is a mainstream product used for everyday activity. In this case, it is a combination of two ‘tech giants’ in Apple Inc. and Google Inc.. Indeed, the numbers involved highlighting the scale of the problem. During the period of this research, Apple made its 2 billionths iOS-based device (Apple Inc., 2018). It is estimated that the number of people over the age of 60 will pass 2 Billion by 2050 (Division, 2017) and age-related issues with near vision, critical for use with the devices explored, is already over a billion (Wolffsohn and Davies, 2018). These numbers are in the context of the unrelenting march of digital technology. The introduction highlighted that in 1986 it was estimated that only 0.8% of the global data storage was digital, by 2007 this had shot up to 94%. The impact of poor UI visibility can be characterised in terms of billions of people interacting with billions of devices in an ever-changing digital landscape. It is important not to drift into hyperbole, but the potential scale of the problem warrants the research that was undertaken and importantly the value of continuing to pursue it. Positive answers to questions 1 and 2 lead to outlining a problem worth solving. The bedrock of the subsequent enquiry were the research questions. Figure 8-2 is a modified final version based on reflections of the process as a whole. These are 'final' in the sense that they represent the result of an iterative approach used to refine them.

Points of note about these questions are, firstly, their hierarchical nature from the broad and practical overarching question to the very specific, which proved useful in the exploration of the problem of UI visibility and looking for potential ways to improve it. Secondly, the first two initial research questions, concerning what UI visibility is and the problems it can cause, were critical in understanding a complex and elusive phenomenon. It can be described as elusive as the naïve model of vision easily dominates thinking, so moving, or more realistically jumping, to one that is cognitively framed is not easy. Vision is complex, as are users more generally, and the user interfaces they interact with.

In this regard, as highlighted above, the research questions and the framework within which they sit provides a good balance of the broad to the specific and the related underlying conceptual understanding required to answer them. The one area, that came out in the latter work with practitioners, is the nature of their current practice. This proved to be highly illuminating and warrants drawing out an additional research question, which is, 'how do practitioners currently evaluate UI visibility?'. This has been added in as SRQ1&2d in Figure 8-2 and is highlighted with two large white arrows. This question will be discussed in more detail later in this section.

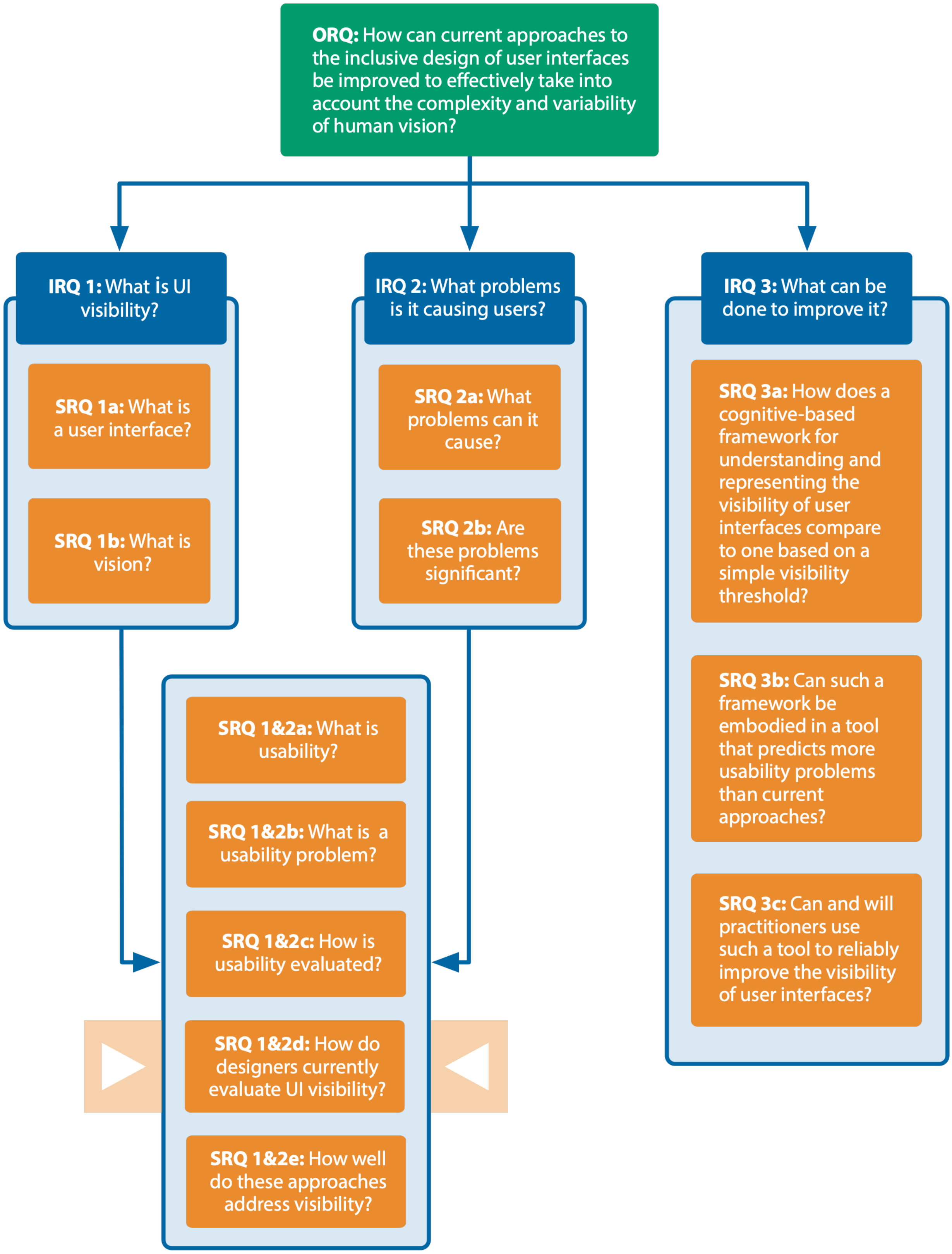


Figure 8-2: Proposed evolution of the research questions (based on Figure 3-28)

These specific research questions were used to structure the literature review and ultimately address the high-level, guiding questions. In this case question number 4, about how well the literature already answers the specific research questions, and question 5 about any gaps that there are in addressing them. A reflection on the literature in light of this research highlights several key points as follows:

- The seemingly obvious nature of UI visibility—its apparent 'can I or can I not see it' property—masks its underlying complexity. This is reflected in the literature where it is not explicitly defined or deconstructed and conceptualised in any detail. This leaves a **gap** of definition and conceptualisation.

- It is also not surprising, that without an underlying conceptualisation, usability evaluation methods do not address visibility from a cognitive perspective. Therefore, there is a **gap** within current approaches opening the potential for them to be augmented and also for completely new evaluation methods.

- The testing of the vis-UI-lise tool confirmed the challenges of defining robustly what usability is and in particular what a usability problem is. These elude definitions that are generalisable and repeatable in practice leading, in part, to the difficulties with interrater reliability in the identification and classification of usability problems. The fundamental nature of casual chains means this is a hard problem to resolve, but none the less there are philosophical and theoretical **gaps** in the literature in this area. This is not only due to the underlying complexity but in part reflects the pragmatic nature of the fields of HCI and inclusive design that have helped deliver real-world benefits through their practical focus. However, there is a real gap here that is beyond this particular research, which would benefit from further clarity and insights in these areas.

- Finally, this naturally leads to a **gap** in the operationalisation of a cognitive approach to UI visibility.

More broadly the gap can be seen as the difference between a naïve model of vision—'seeing is believing' and one based on a cognitive perspective—'believing is seeing'. As already stated, this represents more of a 'jump' than a progression from current thinking.

The first bullet point in the previous list highlights the lack of conceptualisation of UI visibility and this leads to the sixth guiding question about what conceptual understanding is required to study the gaps. The research produced a number of key outputs in this regard as follows:

(a) A UI visibility orientated definition of vision, which is as follows:

> *It is an edge-detecting, dynamic, very slightly delayed, selective, blank-filling, prior-experience-combining, object-inferring, distance, direction and speed-estimating, action-oriented system. It is part of a prolific inference engine, making sense out of an incomplete, noisy, sensory input. Vision is an attentionally-driven, zooming spotlight that outputs a 3D, colour, stabilised, immersive representation of the world that enables effective action within it.*

The key aspect is the attentionally driven spotlight that can zoom in and out. This is in contrast to the 'image in our mind' that is essentially uniform, despite the heterogeneous output across the retina (the 'hill of vision'). This difference between what the eye outputs and what are mind portrays, underpins the naïve model of vision which is being challenged in this work.

(b) With a cognitive-based view of vision combined with an understanding of usability and its measurement, leads to the following definition of UI visibility.

> *UI Visibility concerns the demands the user interface's visual properties put on the user's visual ability in the context of use (task) to produce the desired outcome (goals) in a physical, technical & social environment, where visual ability is considered from a broader cognitive perspective including the user's prior knowledge.*

(c) The notion of a visibility spectrum discussed earlier in the form of the 3 M's Model and shown graphically in Figure 8-3 below.

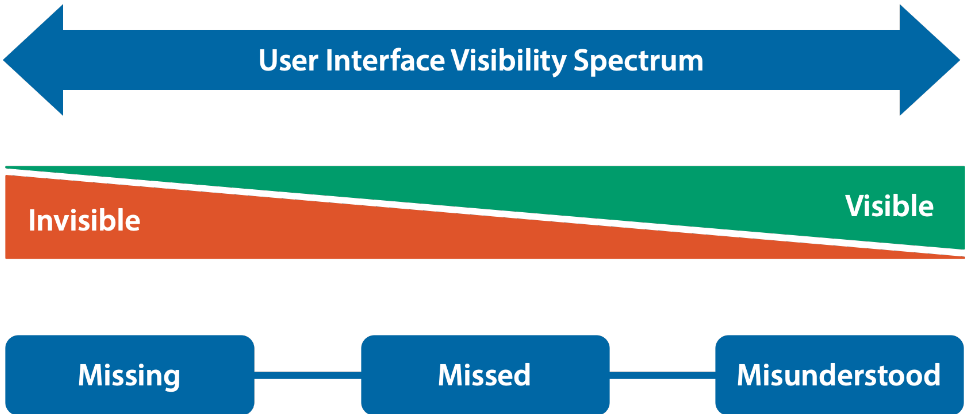


Figure 8-3: 3 M's model showing the visibility spectrum (copy of Figure 3-10)

(d) Combining a cognitive understanding of vison with a model in human-computer interaction led to the 7 P's Model that provides an underpinning conceptual framework and is shown in Figure 8-4. This was heavily inspired by the 3 P's psychophysical model of vision, covering the physics of light, the physiology of the eye and the psychology of the mind (Green, 2008).

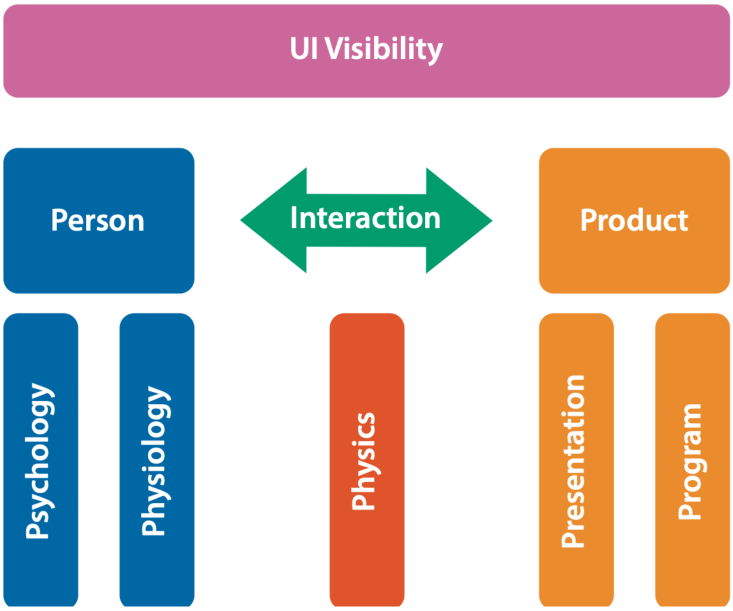


Figure 8-4: 7 P's conceptual framework of UI visibility (copy of Figure 5-2)

(e) The progression to a detailed interaction model, suitable to guide evaluation, was an iterative synthesis across several areas. This included the 3 M's and 7 P's Models previously discussed. Also, it was influenced by existing interaction models and the detailed review of visibility problems from a range of everyday products. However, perhaps the strongest influence in its structure came from the capability-demand model from inclusive design and the associated hurdle metaphor used to structure the approach. This is shown in Figure 8-5. From here the vis-UI-lise tool was developed, key in this is the hurdle questions (See Table 5-5) that drive any evaluation.

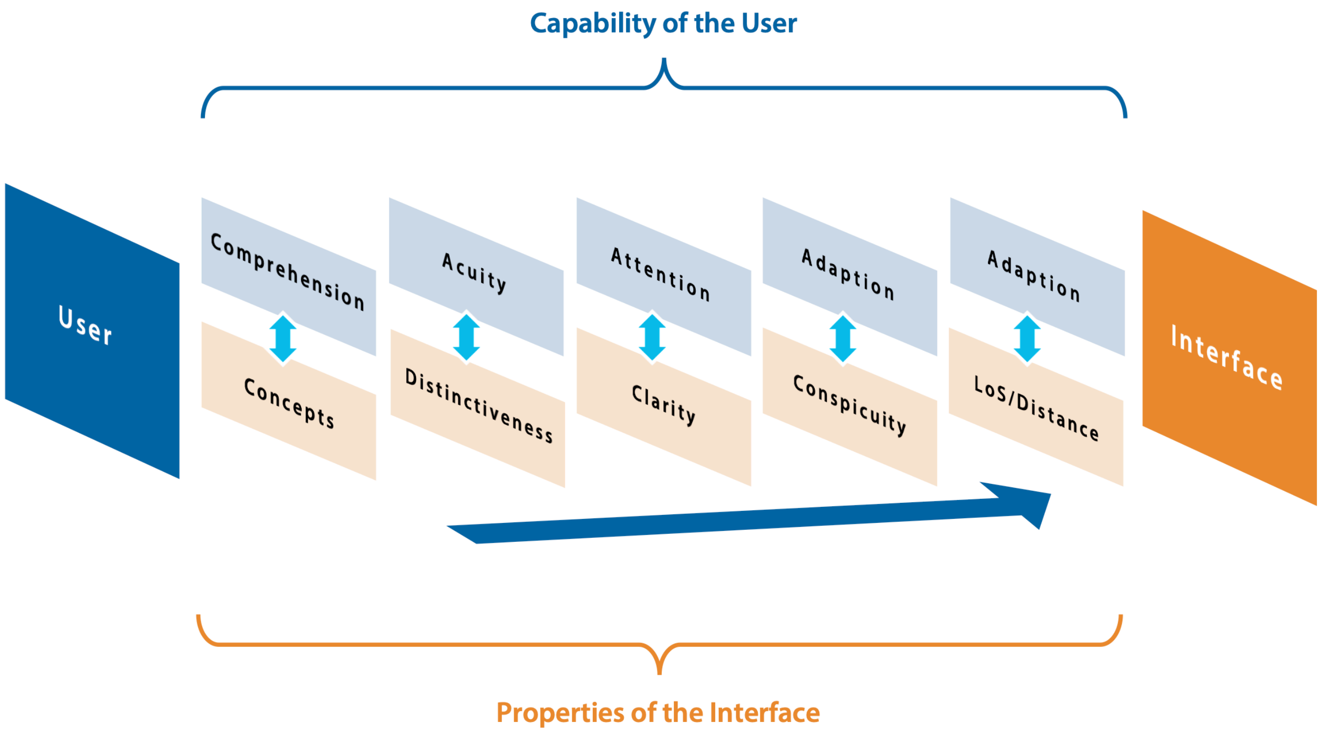


Figure 8-5: Detailed interaction model based on a demand-capability approach (copy of Figure 5-4)

(f) In addition to the conceptualisation described so far, was the development of a detailed generic interaction model, to highlight the range usability metrics available. This was key in determining how UI visibility relates to these measures, in other words, the ways it can impact users and be measured.

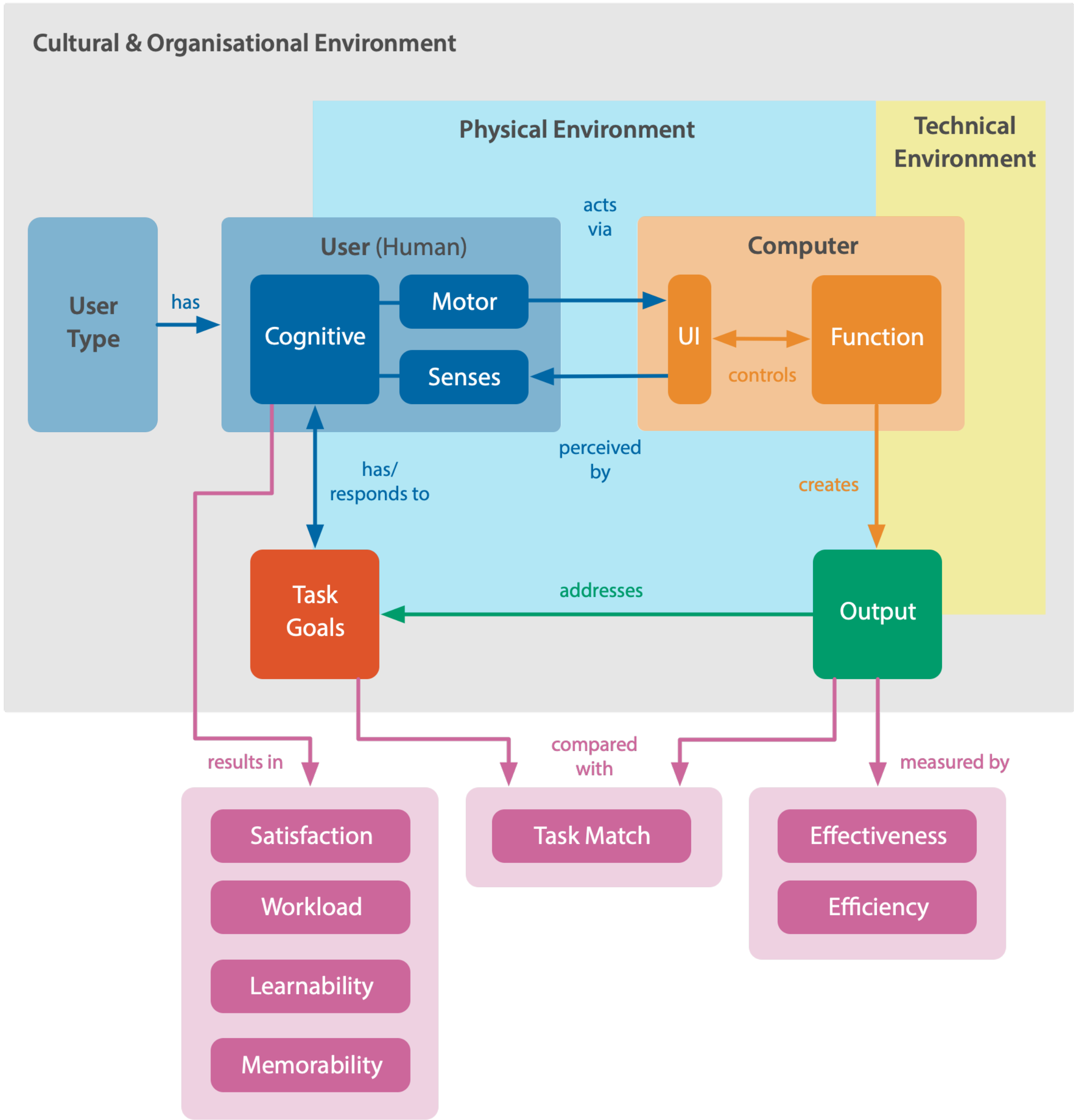


Figure 8-6: An interaction model showing the range of usability metrics that can help measure usability (copy of Figure 3-24)

(g) Another important aspect of interaction is the underlying interface styles. A critique of the historical development of interfaces shows a potential return to the invisibility of command-line interfaces. Table 8-1 below shows this evolution, with two additional descriptions developed for 'touch' and 'gesture' interfaces, based on the existing language for 'command-line' and 'WIMP' interfaces. Such a structure and descriptions are useful for highlighting the context of change that is troubling with the potential degradation in visibility with newer interfaces. In many ways, the WIMP interface typifies the push towards greater visibility to overcome the 'memory' challenges of command-line interfaces. This is now in reverse, particularly for interfaces on small form factor devices. It is an understandable trade-off but there are profound consequences. This description complements the visibility index and is particularly useful in introducing the issue of UI visibility.

Table 8-1: Evolution of interface styles (copy of Table 3-5)

Interface Style	Description
Command Line	Remember and type
WIMP	See and point
Touch interface	Remember and swipe
Gestural	Remember and wave

(h) Finally, the visibility index was a key intermediate step in the development of the thinking around UI visibility. Its limitations were discussed, particularly how it treats all functionality uniformly, regardless of the frequency of use or priority of a function. If all functions were made visible then it is likely to result in visual clutter and the associated challenges with this, therefore user interfaces have to make trade-offs in this regard. None the less, it does provide a useful perspective and complements not only the interface styles descriptions but also the traditional usability measures also outlined above. It also helps in communicating the potential scale of the issue with a particular product, as well as pinpointing specific issues.

Having established a worthwhile problem, with substantive gaps in the literature and a conceptual description of it, it is necessary to have an appropriate empirical approach to research it, which is the seventh high-level, guiding question. In reality, this approach was selected before the work already described and adapted as the research progressed. Figure 8-7 highlights the key elements of the overall research stance. Its graphical form represents the networked nature of the enquiry. This is a key point to make, in that is elaborates on the iterative nature of the approach. Thus, iteration was not a simple repeat of a sequential cycle, but an evolving synthesis (networked) of the different areas that contributed to the development of the thinking and insights gained.

Figure 8-7: Simple model for the key elements of the research approach (copy of Figure 2-1)

This synthesis is best typified by the interaction between the early empirical work and the literature review. This is shown in Figure 8-8 below. Although it is represented as a sequential, iterative cycle it was more of an interaction left to right. As products were explored it drove the review of the literature and the literature, in turn, informed the review of products.

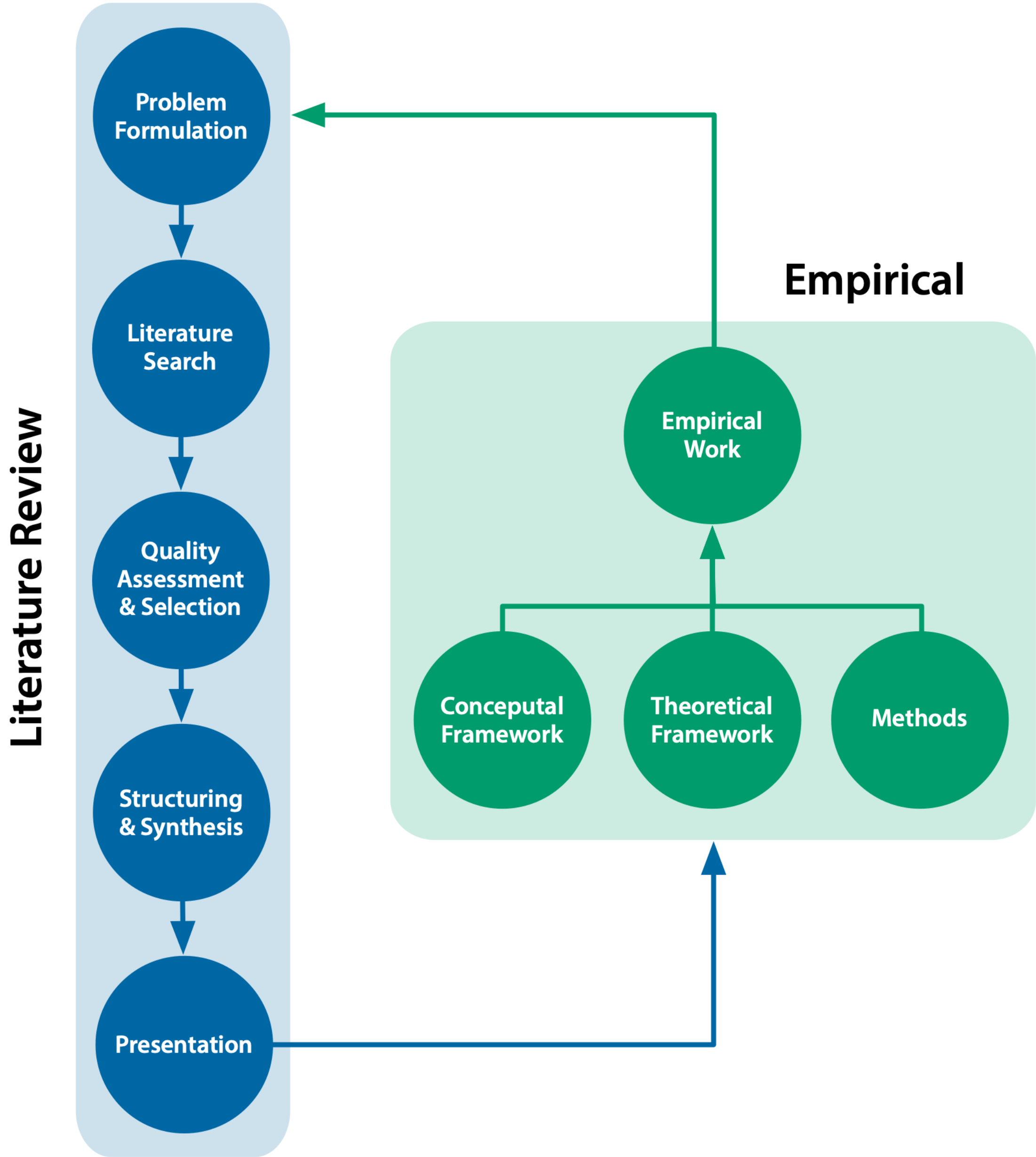


Figure 8-8: Interaction between the literature review and the early empirical work (Copy of Figure 2-6)

The significant advantage of grounding the enquiry in the real examples is that it directly tackles the problems highlighted. The disadvantage is that it limits the scope to what the examples uncovered. This runs the risk of missing significant things that the chosen products do not highlight. For example, none of the examples had issues with the use of colour and the potential problems with colour blindness. This is covered implicitly within the clarity hurdle (see Figure 8-5) and the fourth main hurdle question (see Table 5-5) but is something that would benefit from being made explicit. This is something for future work and could be done, for example, by adding an additional sub-question (4.4) to specifically consider colour use. Despite this limitation, the overall evolutionary synthesis across multiple areas characterised the underlying approach. It is something worth considering when describing an approach as iterative, if it is really a sequential cycle, or as the case here a dynamic and progressive interaction across multiple areas of enquiry.

Regardless of the iterative nature, and the emphasis here on an evolutionary synthesis fed by a network of elements, the research ultimately sits in time, which is linear. The DRM provides a higher-level framework that sits across the broader trajectory of the research and within which the iteration sits. The DRM also espouses iteration, so in this regard, the DRM is a good fit with the approach described. As a framework, it provided a good balance between structure and the freedom within it to adapt to the emergent understanding of UI visibility. The ARC diagram with the addition of annotation with key 'seed' papers proved invaluable at organizing the literature review. The DRM reference model (Figure 3-29) and impact model (Figure 3-30) helped the structure initial thinking and drive the synthesis of multiple inputs described above. An overall methodological view, from the initial motivation to the development of an intervention, is presented graphically in Figure 8-9. This puts the DRM framework and research stance in the wider context of research as a whole, linking to ontology and epistemology and the variety of academic fields that the research drew from.

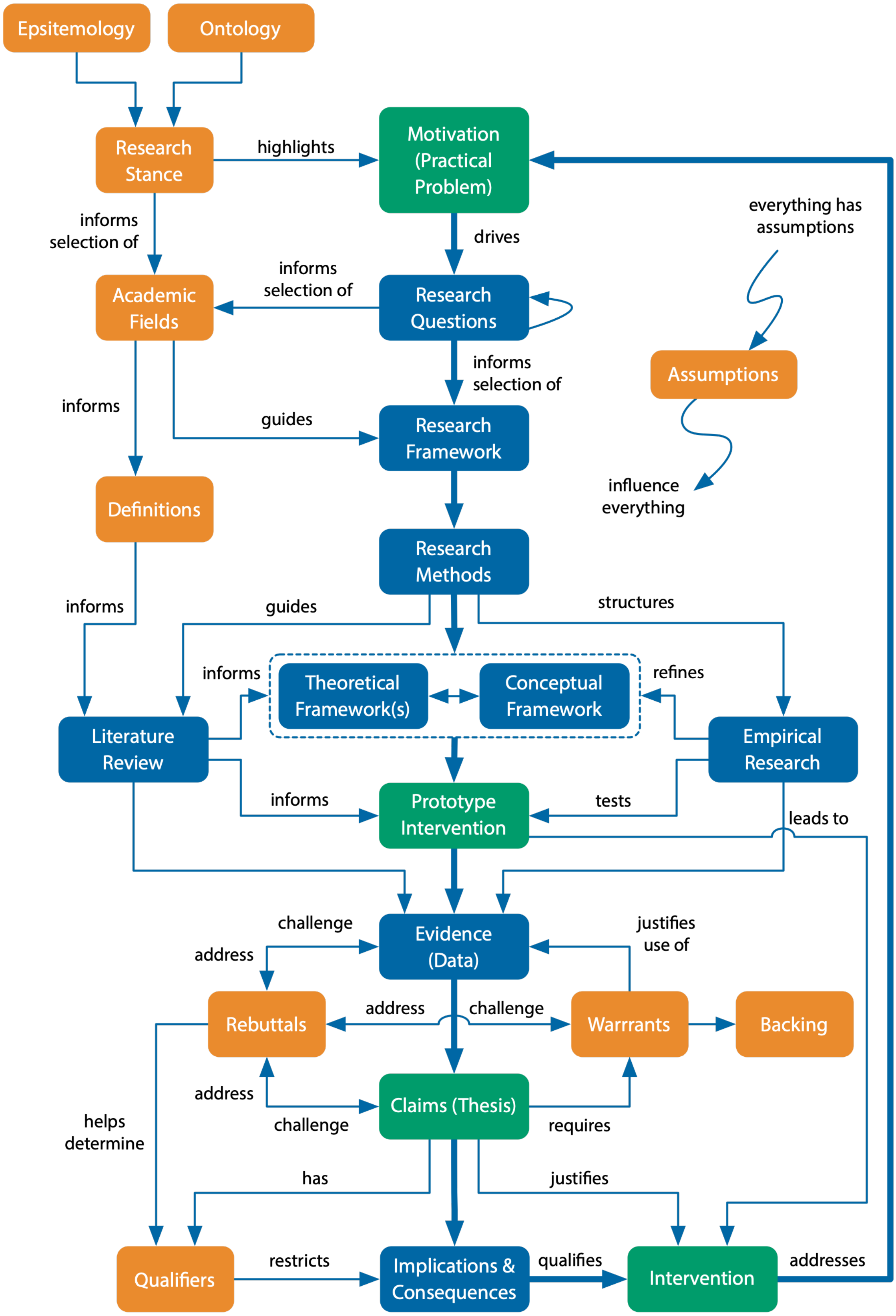


Figure 8-9: The research approach as an end-to-end process (Copy of Figure 2-8)

The research approach outlined led to setting the goal of developing an intervention to help better assess the visibility of user interfaces from a cognitive perspective (psychophysical). This was embodied in a tool called vis-UI-lise, for which a comprehensive support package was developed. This was tested with regard to its predictive power by comparing the predictions it made against those observed with real users. The tool was then tested with practitioners. The results from this testing as well as the preceding work lead to addressing general question 8, which concerns what these results tell us and what claims can be made. This is covered in the following section.

8.2 Research Claims

What follows are a series of claims derived from the research, which relate to the research questions. They can be claims of fact, definition, cause, value, or policy. The claims were developed using the Toulmin argumentation method in conjunction with the different claim types. This approach is described earlier in Section 2.3.1 in Chapter 2.

For each claim, the different aspects are broken down in the constituent parts of the Toulmin model to show the underlying thinking that is brought together in the final stated claim. It starts with a broad, essentially ideal, claim and then moderates it based on the available evidence and potential rebuttals. Where appropriate warrants are stated in the claim if they are not obvious or potentially controversial. This is in line with extensive discussion on warrants and their use by Booth, Colomb and Williams (2008). Within the limitations of the actual wording, the final qualified claim has the initial broad claim in bold and the qualification in regular text to highlight how it has been appropriately modified. The claims are as follows:

Claim 1

Related to: IRQ 1- What is UI Visibility?

Broad claim:

UI Visibility concerns the demands the user interface's visual properties put on the user's visual ability

Claim Type: Definition

Evidence:

This definition is based on the synthesis of four key inputs namely: (1) a psychophysical model of vision; (2) models of human-computer interaction; (3) the demand-capability model from the field of inclusive design (3); and (4) observed problems of visibility with different products.

Warrant:

It is assumed that new definitions can be derived by combining input from multiple sources.

Backing:

The warrant is a broad appeal to the academic process of the synthesis of different sources. The definition is indirectly tested through the empirical work.

Rebuttal:

The definition narrows itself to a cognitive view of vision and it could be argued that it should be framed independently of a particular model of vision. The rationale for doing this is that without making the cognitive dimension explicit it does not cover the visibility spectrum from 'missing' to 'misunderstood'. The broad claim is devoid of the context that product use sits within, which is deemed to have a large influence on usability outcomes.

Qualifier:

This definition represents a starting point for further research. As such it will benefit from scrutiny and further refinement.

Claim 1:

UI Visibility concerns the demands the user interface's visual properties put on the user's visual ability in the context of use (task) to produce the desired outcome (goals) in a physical, technical & social environment, where visual ability is considered from a broader cognitive perspective including the user's prior knowledge.

Claim 2

Related to: IRQ 2- What problems is it causing users? and SRQ 2a - What problems can it cause?

Broad claim:

There are three types of UI visibility problems. Firstly, some user interface elements have no visible component at all, from a visibility perspective they are effectively 'missing'. Secondly, if an element is present then they can be 'missed'. Thirdly, if they are seen they may be 'misunderstood' by the user.

Claim Type: Definition and to a degree cause

Evidence:

The categorisation of problem types is based on their identification within actual examples of different products and the observation of users.

Warrant:

Direct observation of issues is deemed to be a strong form of evidence.

Backing:

The high value placed on empirical evidence is well established.

Rebuttal:

The sample of products was limited and therefore the categorisation may not be sufficient to cover all problem types. Problems may also not fit neatly into these categories.

Qualifier:

This is only one possible model and issues may not fit into the three categories or cover all instances of problems. These problems may map to higher-order usability problems in different ways e.g. effectiveness, efficiency and satisfaction.

Claim 2:

UI visibility problems can be broken down into at least **three types. Firstly, some user interface elements have no visible component at all, from a visibility perspective they are effectively 'missing'. Secondly, if an element is present then they can be 'missed'. Thirdly, if they are seen they may be 'misunderstood' by the user.** This is based on the detailed analysis of a range of different products and as such represent the direct observation of issues and therefore has an empirical basis. Although not necessarily exhaustive to all problem types, it provides a useful categorization of the different types of UI visibility problems a user may encounter and that in turn may result in poor user experience.

Claim 3

Related to: SRQ 2a - Are these problems significant?

Broad claim:

The shift to and widespread use of mobile devices with gesture-based interfaces (touch) has resulted in many user interfaces having lower visibility.

Claim Type: Value

Evidence:

Apple alone has sold over 2 billion devices based on the mobile operating system iOS. These devices have touch interfaces with constraints in the size of the screen which can lead to user interface elements being 'invisible' to avoid visual clutter. Analysis of an iPhone 7 showed the 'visibility index' for just navigating around the various home screens was only 8% of the 622 functions available. A less detailed analysis of similar functionality of the Apple Macintosh™ (MacOS) from 1988 (System 6) and 2013 (Mac OX 10.9) showed figures of 91% and 63% respectively. Testing with real users using a phone for an everyday task (emailing a selfie) showed that 81% of unique problems observed were visibility related.

Warrant:

A combination of different data is used to deduce a potential trend.

Backing:

There is a logical link between population numbers, devices and the visibility of different interface styles.

Rebuttal:

The WIMP benchmark figures are not directly comparable in terms of functionality. All the functionality considered is very restricted. Regardless of the trend, it is plausible that people will adapt to new interface thus mitigating the visibility issues.

Qualifier:

The trend can be described as 'likely' as opposed to a more certain form of language. The issue of users potentially adapting to the trend, thus negating its impact, is addressed in terms of the problem of older users who are more likely to struggle to adapt.

Claim 3:

The shift to and widespread use of mobile devices with gesture-based interfaces (touch) is likely to have **resulted in many user interfaces having lower visibility** compared to traditional WIMP based interfaces. Apple Inc. alone has sold over 2 billion devices based on their mobile operating system (iOS) which impacts a large proportion of the global population. These devices have touch interfaces that typically have constrained screen sizes that rely on the use of gestures instead of buttons and a pointer. An analysis of an iPhone 7 (running iOS 10) that just considered the basic navigation around apps in the home screens showed that only 8% of the 622 functions (section 4.3) were visible at the top level. A reading of a manual for such devices will highlight the range of gestures that have little or no visual indication of their availability to the user. Apple Inc. has even removed the familiar home button since the iPhone X onwards. This compares to older WIMP interfaces were a less detailed analysis of similar functionality of the Apple Macintosh™ (MacOS) from 1988 (System 6) and 2013 (Mac OX 10.9) showed much higher figures of 91% and 63% respectively. In testing with users performing an everyday task on a Windows Mobile™ phone showed that 81% of unique problems observed were visibility related (section 6.6). Although users, may to a degree, memorise 'invisible' gestures it is of particular concern for older users who are likely to struggle to learn them.

Claim 4

Related to: SRQ 1&2d – How well do these approaches (current UEMs) address visibility?

Broad claim:
Current UEMs, where visibility is explicitly addressed, only address it from a simple threshold perspective and as such they cannot address the range of visibility issues as represented by the 3 M's model.

Claim Type: Value

Evidence:
A review of 8 different perceived (subjective) usability instruments (Table 3-9), 14 broad usability approaches (Table 3-10) and 12 human factors standards (Table 3-10) showed that none of the approaches directly addressed visibility from a broader cognitive perspective.

Warrant:
A range of evaluation approaches was directly assessed with regards to visibility.

Backing:
A systematic literature review was undertaken to establish the predominant approaches described in the literature.

Rebuttal:
Current UEMs and related approaches may not address visibility from a broader cognitive perspective as it makes little or no difference in identifying usability problems. The cognitive dimension may also be addressed by direct consideration of cognitive factors. Also, UEMs may have been missed that do take a broader cognitive approach to visibility.

Qualifier:
Make explicit the number of UEMs and related approaches that were reviewed and acknowledge that approaches may tackle issues in different ways.

Claim 4:

A review of 34 different evaluation approaches covering perceived usability measurement, usability evaluation methods and human factors standards showed that **where visibility is explicitly addressed, it only addresses it from a simple threshold perspective. As such they cannot** directly **address the range of visibility issues as represented by the 3 M's model,** although they may be tackled indirectly via other means.

Claim 5

Related to: IRQ 3: How can it be improved? and SRQ Question 3a: Is a cognitive-based framework for understanding and representing the visibility of user interfaces more effective than one based on a simple visibility threshold?

Broad claim:
The 7 P's Conceptual Model (Figure 8-4) and detailed interaction model (Figure 8-5) based on it represent a more effective description of UI visibility than one based on a simple visibility threshold.

Claim Type: Value

Evidence:
The model is based on a more complete model of vision. As such it can explain observed problems such as functions that are 'missing', 'missed' or

'misunderstood'. It has subsequently been used to develop the vis-UI-lise tool that has shown promise at predicting problems in initial testing.

Warrant:

A more complete understanding of the underlying contributory factors, in this case, vision, is likely to lead to a better understanding overall.

Backing:

A simple model of vision does not address observed problems of the 3 M's or other comparable phenomena such as inattentional blindness.

Rebuttal:

It could be possible to combine a simple threshold model of vision with a general cognitive view rather than integrate it. The proposed model is complex that may not lead to greater insights in practice. There is a lack of empirical evidence that the approach is more effective.

Qualifier:

The comparison is made in terms of 'completeness' rather than 'effectiveness'. It is positioned as a starting point for critique and iteration.

Claim 5:

The 7 P's Conceptual Model (Figure 8-4) and detailed interaction model (Figure 8-5) based on it offers a more complete **description of UI visibility than one based on a simple visibility threshold.** The framework is specifically able to identify problems of functions that are 'missing', 'missed' and 'misunderstood'. The detailed interaction model formed the basis of the vis-UI-lise evaluation tool that has shown significant promise in initial testing. Although the research is at an early stage it is a promising platform on which to base further work.

Claim 6

Related to: SRQ 3b: Can such a framework be embodied in an evaluation tool that predicts more usability problems than current approaches?

Broad claim:
The vis-UI-lise evaluation tool was successfully based on a cognitive model of vision and performed well compared to existing evaluation methods.

Claim Type: Fact and Value

Evidence:
On a range of UEM specific performance scores the vis-UI-lise tool performed as well as or better than many of the established UEM tools.

Warrant:
The UEM scores are valid across different types of UEMs.

Backing:
The scores are agreed amongst experts and documented in the literature as a way of evaluating different UEMs

Rebuttal:
A direct assessment of vis-UI-lise against different UEMs against a common problem set was not made. The vis-UI-lise scores were for the developer of the tool who is also a very experienced evaluator.

Qualifier:
It relates to initial testing. The comparison with other UEMs is to a range of typical scores. Therefore, it shows potential.

Claim 6:

The vis-UI-lise evaluation tool was successfully based on a cognitive model of vision and in initial testing **performed well compared to** typical scores **for existing evaluation methods (UEMs).** vis-UI-lise predicted 85 out of the 113 unique problems encountered by participants during user testing. This gives a thoroughness score of 75% that exceeds the majority of the scores of other UEMs it was compared to that are documented in the literature. However, its validity score is low at 41% and the meaning of this is hard to assess as feasibly the problems that were predicted may have a low probability of occurring and would not necessarily show up with the sample size of 14 users. The lack of a known 'complete' reference set of problems is a fundamental problem with trying to score any UEM. Overall the scores show that the vis-UI-lise tool has great promise in terms of its potential performance, but as with all UEMs testing in different contexts is critical to gaining a more complete picture of how it can perform.

Claim 7:

Related to: SRQ 3b: Can and will designers use such a tool to reliably improve the visibility of user interfaces?

Broad claim:
Usability practitioners were positive about the rigour that the vis-UI-lise tool offers in evaluating product interfaces

Claim Type: Value

Evidence (data):

A study with four different companies using the tool in various forms led to follow-up interviews that were transcribed and analysed. The comments about rigour were spontaneous from the practitioners and not as a result of specific prompting with regard to rigour.

Warrant:

Direct feedback from users of the product is a valued form of evaluation.

Backing:

'Eyewitness' accounts are valued as a form of evidence.

Rebuttal:

The sample size is small (4 companies) and no one used the tool consistently over an extended period. Despite practitioners being positive about the rigour, they experienced problems with the complexity of the tool.

Qualifier:

Restrict the claim to it being based on initial testing. Contrast the positive aspect of rigour with the issue of the tool's complexity. Limit the scope of the claim to be the justification for further work.

Claim 7:

Initial testing with usability practitioners showed that they liked **the rigour of the vis-UI-lise tool** but struggled with the overall complexity of language used to describe the various elements. As such it represents a strong base from which to undertake further development to draw out the strengths and mitigate the weaknesses in its current embodiment.

Claim 8

Related to: SRQ 3c: Can and will practitioners use such a tool to reliably improve the visibility of user interfaces?

Broad claim:
Usability evaluation tools need to be flexible enough to adapt to different project contexts.

Claim Type: Fact

Evidence:
Interviews with usability practitioners about a range of projects they have been involved in.

Warrant:
Direct feedback from practitioners about their own experiences.

Backing:
'Eyewitness' accounts are valued as a form of evidence.

Rebuttal:
The sample size is limited, and it is an interpretation of what practitioners said. The sample may not be representative of different types of projects.

Qualifier:
Limit the claim to initial testing. Direct its application to justify further investigation to help improve the vis-UI-lise tool.

Claim 8:

Initial structured interviews with usability practitioners highlighted that, with regard to usability evaluation tools, 'one size' does not fit all. **They expressed the need to significantly adapt any usability evaluation approach to the context** it is being used in to address issues such as client preferences, products specific issues and resources limitations such as time and budget. This is a key area that requires detailed further investigation to help drive improvements to vis-UI-lise.

Claim 9

Related to: SRQ 3c: Can and will practitioners use such a tool to reliably improve the visibility of user interfaces?

Broad claim:
Five areas have been identified for further development based on the current research and the vis-UI-lise tool. These are: (1) make further refinements to the vis-UI-lise tool; (2) produce a specific variant of the vis-UI-lise tool for medical device development; (3) augment Nielsen's 10 heuristics to take greater consideration of UI visibility; (4) develop simple awareness-raising tools such as simulation of the zooming spotlight of vision; and (5) develop the visibility index approach and determine how it can best be applied.

Claim Type: Fact and Policy

Evidence (data):
Interviews with usability practitioners about the vis-UI-lise tool and their current approaches led to reflecting on a range of possible developments.

Warrant:
Direct input from users of the product is a valued form of evaluation.

Backing:
Stakeholder involvement is deemed to be good practice to ensure interventions are fit for purpose.

Rebuttal:
The suggested developments are based on a limited sample size and the suggestions have not been validated with practitioners.

Qualifier:
Limit the claim to "potential areas of development" rather than it being a definitive and validated list.

Claim 9:

From discussion with usability practitioners and reflection on the research as a whole has helped **identify five** potential **areas of development for improving on the research undertaken to date. These are: (1) make further refinements to the vis-UI-lise tool; (2) produce a specific variant of the vis-UI-lise tool for medical device development; (3) augment Nielsen's 10 heuristics to take greater consideration of UI visibility; (4) develop simple awareness-raising tools such as simulation of the zooming spotlight of vision; and (5) develop the visibility index approach and determine how it can best be applied.**

Claim 10

Related to: ORA: How can current approaches to the inclusive design of user interfaces be improved to more effectively address the complexity and variability of human vision?

Broad claim:
More research in taking a cognitive perspective of vision with regards to UI visibility is a priority, as is the development of evaluation tools to accompany it.

Claim Type: Policy

Evidence:
UI visibility is key and UI visibility is likely to be getting worse. Current approaches do not address the different aspects of visibility problems as described in the 3 M's model. vis-UI-lise shows a practical approach to addressing the broader issues of visibility.

Warrant:
Research is key to help drive an iterative process to improvement.

Backing:
Iteration and evaluation within it are well-established parts of improvement processes.

Rebuttal:
Claiming something is a priority implies and requires evaluation against either potential interventions.

Qualifier:
Changing the wording from a 'is a priority' to 'should be considered as a priority' on the basis that UI visibility is a core component of usability and will inherently be something that should be considered regardless of other competing priorities.

Claim 10:

More research and development into UI visibility should be considered **as a priority due to the scale of the problem,** its impact on usability and indications that it is getting worse with the trends in mobile and touch devices. Work to date has shown the potential of novel approaches such as the vis-UI-lise evaluation tool, but more work is required to address current weaknesses and build on the strengths shown.

8.3 Academic Contribution

The answers given to the high-level guiding questions 1 to 8 (Section 8.1) and the ten claims (Section 8.2) represent a summary of the academic contributions. These contributions are the topic of high-level question 9 and are mapped to the research questions in Figure 8-10, which is split across the following two pages. The map also includes: the ten claims; the underlying empirical work; as well as three papers resulting directly from this research, which were published as part of conference proceedings.

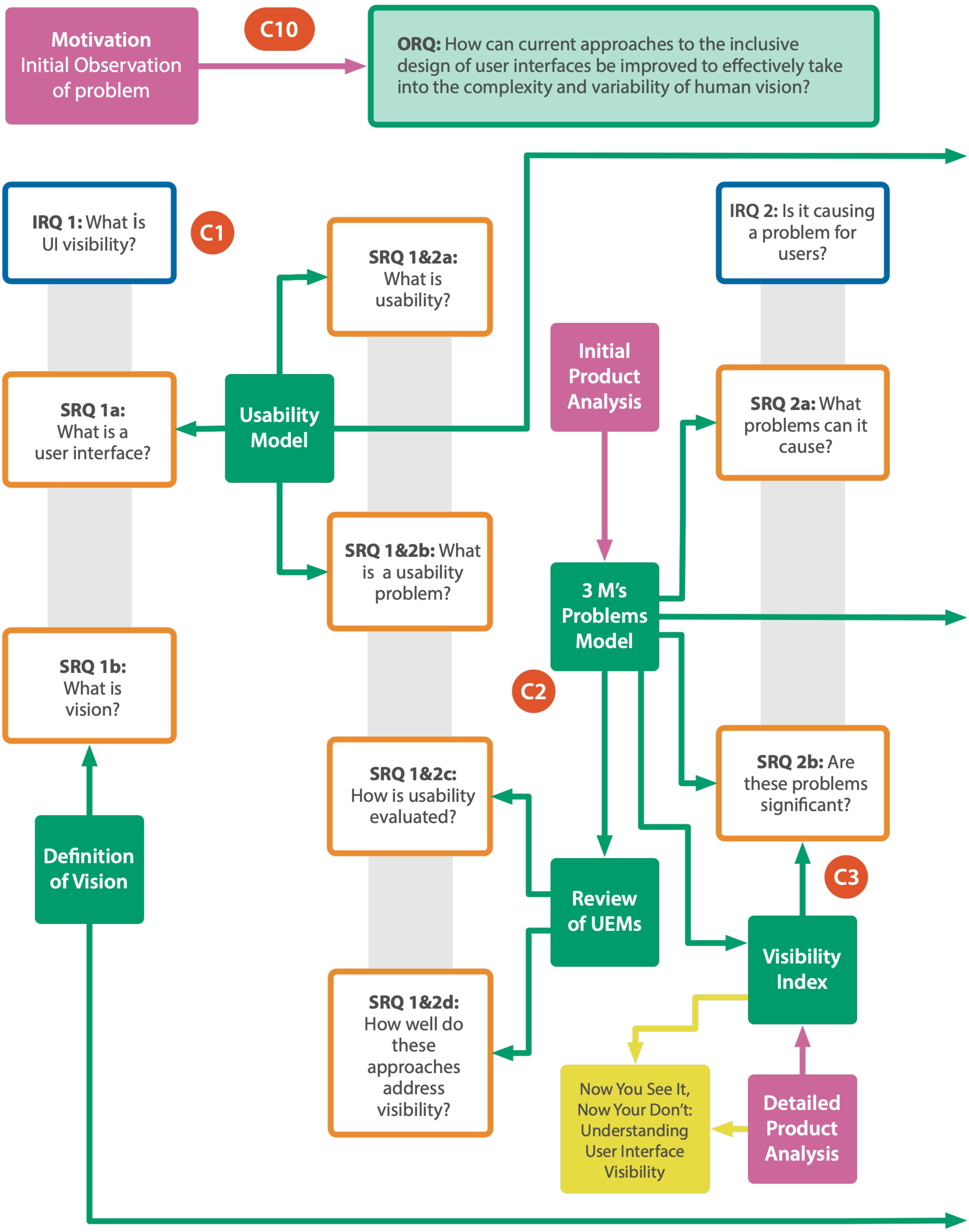


Figure 8-10: Mapping of key research and contributions to the research questions

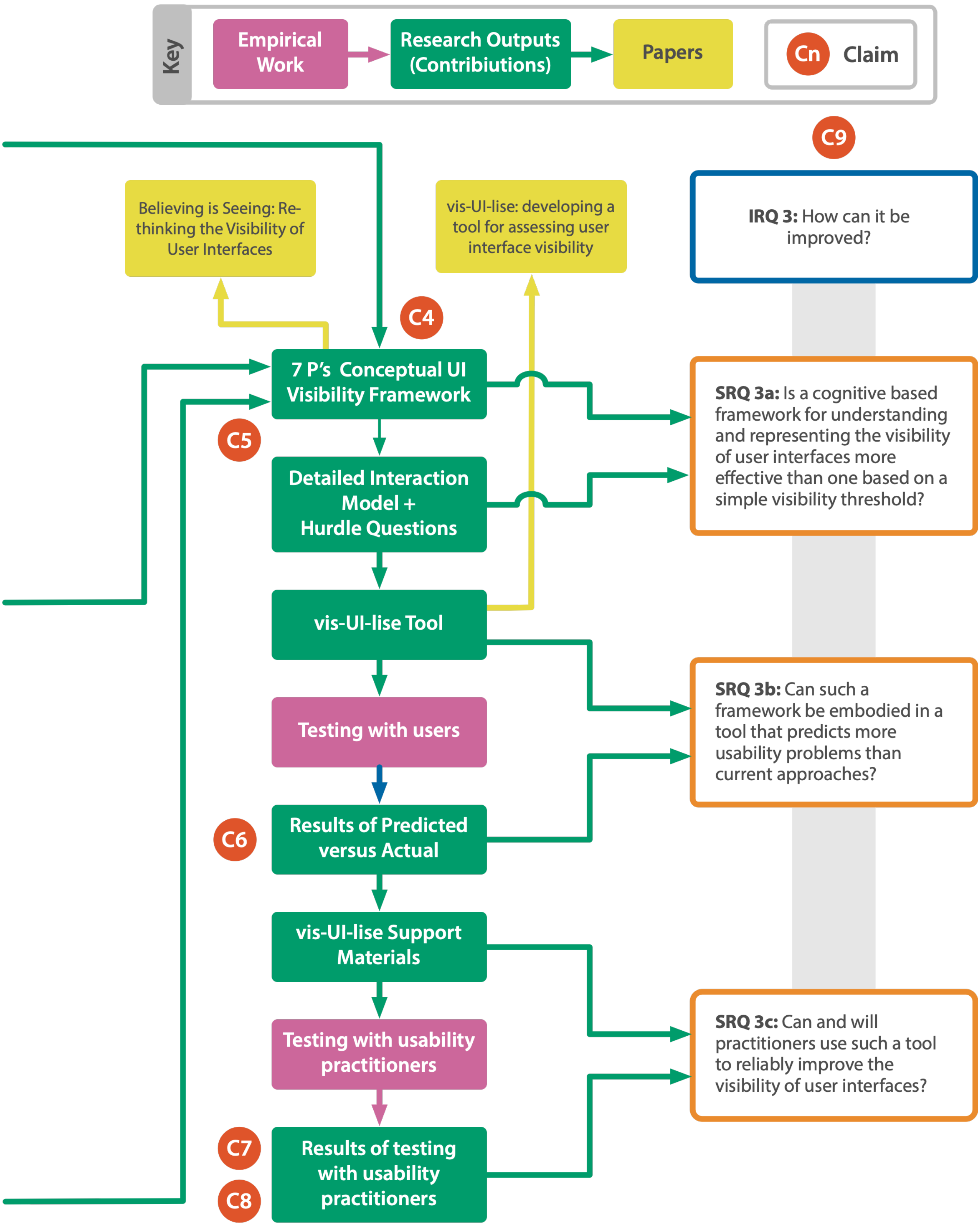


Figure 8-10: Continued

8.4 Limitations & Issues

Limitations and issues were identified and addressed throughout the research and covered directly in the thesis concerning specific elements of the research. This section summarises the key areas of limitation to guide appropriate use of the outputs. These limitations are divided into ten areas as follows.

Firstly, and a key foundational challenge, are difficulties surrounding the **definitions** of 'usability' and 'usability problems'. These were discussed in Sections 3.7 and 3.8 respectively. Usability is not easy to define, multidimensional, and context-dependent. This ultimately limits what can be said about the efficacy of any intervention, such as the vis-UI-lise tool, to improve it. Put simply, if it is hard to define it is also hard to say if you have made things better, which was the key driver for the research. A pragmatic approach was taken to work within the current definitions with the acknowledged limitations. Further mitigation can be taken through methodological diversity, around evaluation in particular, which will be discussed in point five.

Secondly, the identification of UI visibility problems was limited by **product set size and selection** (diversity) of products. This was due to the inevitable constraint of resources (time and money). Put simply, will the product set highlight the range of visibility problems that exist. Section 3.2.1 discusses the rationale of selection around targeting products covering different instrumental activities for daily living (IADLs) leading to product category diversity. Also, coverage of different forms of UIs from small to large was considered, as well as, to a limited extent, UI evolution over time. This selection framework could be used in future work to help systematically increase product diversity.

Thirdly, the testing of the vis-UI-lise was constrained by a range of well-established **issues with UEM evaluation**. These are discussed in detail in Section 6.2.1. but the high-level issues of note are: (1) the problems that result from difficulties with definitions; (2) the lack of a direct benchmark comparison with other UEM tools; (3) the known problem of 'evaluator effect', in other words, the variability of output from a UEM tool across different evaluators; (4) and the decontextualized nature of a laboratory-style evaluation. The testing made use of different UEM performance metrics and a range of UEM test results that are in the literature. This represents a good starting point to assess the potential of the vis-UI-lise tool. If greater resources were available, then testing with more evaluators and directly against established UEMs would address some of these issues. However, regardless of any resource constraints, for the first iteration of the tool the approach taken seems appropriate to be able to justify further work.

Fourthly, **evaluator** or more specifically **'creator' bias** is an inevitable risk in evaluating your own work. This route was taken in part to ensure that all the work was that of the author, in line with the requirements of a PhD. However, regardless of this, there is value in evaluating your own work, in that the creator has a deeper understanding of the content and therefore attuned to potential issues. Aligned with this, the stance taken in the evaluation was one of problem finding and improvement. This framing moves things away from trying to show that the vis-UI-lise tool is 'better' than other UEM's. Going forward independent review would be beneficial, but again at this early stage, the approach taken is not inappropriate.

Fifthly, there is a **lack of methodological diversity**. Again, this is down to deliberate choices in line with the stage of the research and available resources. The lab-style approach to the testing of the vis-UI-lise tool is particularly narrow and does not address likely contextual differences in real-world use. This was a trade-off to be able to make a comparison with UEM performance measures from other published research as discussed in point three above. Future work should look to complement the quantitative approaches with more qualitative evaluation and where possible the use of vis-UI-lise by usability practitioners on real projects.

Sixthly, and linked to methodological diversity, is the need for **repeated experiments** to address the range of issues already outlined. Part of the aim of this phase of research was to develop a justification for further research, in other words, test it to a level and standard to produce convincing evidence that warrants and encourages further investigation.

Seventhly, the **role of culture** in visibility is noticeable by its absence in much of the literature that was reviewed. This finding is consistent with Masuda (2010, p. 339) who highlights that mainstream psychology assumes that perceptual processes are *"universal"* and driven by the "underlying optical mechanisms and characteristics of visual information hardwired in the human brain and shared by human beings in general". This perspective is now being challenged leading to a view that culture and human psychological processes interact with each other. The reference to the role of language shaping colour perception in Chapter 1 indicates the potential significance of this with regard to UI visibility. This thesis has demonstrated the importance of the cognitive dimension and therefore there is a strong case to extend this further to the cultural dimension too.

Eighthly, there was a lack of **expert input across the domains**, this is particularly pertinent for this endeavour due to its broad multidisciplinary nature as outline in Section 2.2.2. This is justified because it is not possible to form a diverse project team for a PhD. Although, it should be noted that advice was sought during the work from different experts. This phase of the work has created the necessary framing of the problem to be able to engage experts from other domains in the future. Areas of input that would be valuable are: vision science to cover the fundamentals of visual perception; applied psychology to address the broader cognitive issues of product interaction and sociology to understand the adoption by usability practitioners in their work.

Ninthly, there is a **lack of a deeper understanding** of **current design and evaluation practice**. Chapter 7 addresses this in part through the evaluation of vis-UI-lise with usability practitioners. However, it did reveal the complexity, variability and nuances of evaluation approaches. It could be argued this work should have been done at an earlier stage to guide the vis-UI-lise tool. However, it is the conundrum that the co-evolution of the problem and solution space brings (see Section 7.6.4). In other words, having a solution helps to further illuminate the problem. On reflection, having the vis-UI-lise tool and the associated conceptual understanding helped not only frame the dialogue with practitioners but also stimulate valuable insights. Therefore, on balance, the timing was probably appropriate but more work is required in this area.

Tenthly, and finally, there is a temptation to **over generalise the results** into a wider context. The results for the vis-UI-lise tool are very encouraging, but the issues and limitations outlined above show that caution is required. The Toulmin argumentation approach was adopted to ensure that the claims made in Section 8.2 are suitably qualified in this regard and if read in isolation of the research the claims state the limitations of the evidence and guide how they should be applied.

The discussion of these key areas of limitation has pointed towards future work and the next section describes specific recommendations on how to develop the work further.

8.5 Future Work

Seven recommended areas of further work are selected as priorities from the research. The first four come from the evaluation of the vis-UI-lise tool with usability practitioners as follows: (1) refining the hurdle questions (see Section 7.6.1 for suggestions) is key as they are the core of the vis-UI-lise tool in operation. This should also include giving consideration to expanding the questions to explicitly evaluate the use of colour (potentially an additional question 4.4 in Table 5-5). Section 7.6.3 highlights two specific opportunities for variants of the tool. These are: (2) to augment the widely used Nielsen and Mack's (1994) usability heuristics with specific additions around UI visibility; and (3) to integrate the vis-UI-lise tool with the PCA (Perception-Cognition-Action) model and uFEMA (user Failure Effects and Mode Analysis) for medical device evaluation. The latter is a particular domain that values rigour highly, due to the safety-critical nature of medical device use and the tight regulatory controls around device certification. An alternative to the vis-UI-ise tool was proposed leading to: (4) a 'vision spotlight' simulator. This would help raise awareness of the nature of vision, as well as potentially help determine interface elements that may be 'missed'.

In addition to suggestions that come from the evaluation with practitioners there are several other interesting additional areas worthy of further investigation. The first of these is; (5) the potential in the visibility index described in Chapter 4. A good starting point for this would be to see if the visibility index correlates with usability performance, in other words, if a low index value correlates with poorer usability performance. If this is the case, then the index could prove to be a powerful proxy measure for usability. Related to this is; (6) existing visual clutter metrics that are outlined towards the end of Section 3.5. These are more abstracted than the visibility index, but again could be a good proxy measure, and could be very quick to apply as the user interface is evaluated by software-based analysis of the images of the interface. Finally, (7) extending the research boundaries from the cognitive perspective to wider cultural implications on UI visibility as outlined in the previous section on limitations and issues.

Earlier in Section 5.6.1 the case for creating a standalone UEM was made to enable the focus on UI visibility to be unencumbered by current approaches. The recommendations outlined above show the evolution from this point to three broad categories for positioning the work going forward. The first, continues the development of **standalone** tools and approaches to address UI visibility, as exemplified by recommendations 1,4,5, and 6. The potential for later **integration** into existing UEMs, set out in Section 5.6.1, is realised in the recommendations regarding integration into two well established UEM approaches (recommendations 2,3). Finally, the hurdle questions (recommendation 1) and underpinning models can be used to **inform** the development of both standalone and integrated tools and indeed inform the wider discourse in UI visibility.

These seven recommendations and the three broad areas of positioning are pragmatic in their focus and could lead to relatively simple interventions which could transform current practice.

8.6 Methodological reflections

There are a number of things, from a personal perspective, that stand out with regards to methodology and more generally to my own practice. Firstly, the Toulmin argumentation model is not the easiest to grapple with but proved to be a powerful approach for improving the qualification and rationale for making claims. Secondly, the work has resulted in an ever-growing appreciation of qualitative work. This particularly applies to a Charmazian grounded approach. Thirdly, it has resulted in a further shift from a positivist to a more interpretivist research perspective. This has also reinforced a rooting in a critical realist stance that affords drawing on and valuing different methods.

Fourthly, concerning my work as a usability practitioner I now look at user interfaces differently, considering how UI elements can be 'missing', 'missed' or 'misunderstood'. At a more profound level is a shift to placing a greater value on qualitative findings as mentioned above. The 'numbers' in quantitative work can exert an often-unwarranted authority that qualitative work would 'struggle' to do. I now treat qualitative and quantitative evidence differently when making decisions in projects, in particular having greater caution when reviewing quantitative findings. Additionally, I am also much more cautious in the generalisation of results, particularly from single studies. I now take much greater consideration of the context of the original research and how applicable it might be in a different context. Ultimately projects require decisions to be made and risks to be weighed up based on the available evidence. I hope that my decision making has been improved by the lessons I have learnt from this and other related research.

Finally, the words of G.K. Chesterton (1929) summarise the respect that has grown for what has gone before.

> *"In the matter of reforming things, as distinct from deforming them, there is one plain and simple principle; a principle which will probably be called a paradox. There exists in such a case a certain institution or law; let us say, for the sake of simplicity, a fence or gate erected across a road. The more modern type of reformer goes gaily up to it and says, "I don't see the use of this; let us clear it away." To which the more intelligent type of reformer will do well to answer: "If you don't see the use of it, I certainly won't let you clear it away. Go away and think. Then, when you can come back and tell me that you do see the use of it, I may allow you to destroy it."*

I hope this research is about "reforming" and not "deforming". To stretch the fence metaphor, I have not "cleared away" but built on a body of research to further improve what has already made a large impact on our everyday lives.

8.7 Conclusion

UI visibility represents a problem of global scale that impacts almost everyone's lives. Billions of people are interacting with billions of devices every day to participate in the ever-growing digital world. These interactions are often critical to engaging with society and the visibility of the user interfaces that enable it are typically key in this interaction being successful. The indications are that UI visibility is getting worse, even though our reliance on these devices is increasing.

To compound this UI visibility is complex, there is more to it than 'meets the eye'. Simple threshold-based models based on the premise that 'seeing is believing' are inadequate to understand and assess the visibility of user interfaces where functions go 'missing' or are 'misunderstood'. The cognitive dimension–that in fact 'believing is seeing–is critical in understanding and evaluating what is going on. This research has produced a more rigorous cognitive-based understanding of what UI visibility is resulting in a definition and associated models. This represents the foundation for developing interventions that will help improve current practice. Returning to the overarching research question of:

How can current approaches to the inclusive design of user interfaces be improved to more effectively address the complexity and variability of human vision?

This has been answered through the development of the vis-UI-lise tool that has shown strong promise in initial testing in comparison to existing approaches. In addition, there are concrete proposals for further improvements through: (1) customising the vis-UI-lise tool for use in medical device development; (2) augmenting the widely used usability heuristics with additions addressing UI visibility; and finally (3) developing a vision spotlight simulator. These outputs represent an academically grounded, pragmatic and significant contribution to the problem of UI visibility.

REFERENCES

Abran, A., Khelifi, A., Suryn, W. and Seffah, A., 2003. Usability meanings and interpretations in ISO standards. *Software Quality Journal*, 11(4), pp.325–338.

Adams, W.J., Graf, E.W. and Ernst, M.O., 2004. Experience can change the 'light-from-above' prior. *Nature Neuroscience*, 7(10), pp.1057–1058.

Anderson, J.R. and Lebiere, C.J., 1998. *The Atomic Components of Thought*. Lawrence Erlbaum Associates.

Andre, T.S., Hartson, H.R., Belz, S.M. and McCreary, F.A., 2001. The user action framework: a reliable foundation for usability engineering support tools. *Int. J. Hum.-Comput. Stud. ()*, 54(1), pp.107–136.

Annett, J. and Duncan, K.D., 1967. Task Analysis and Training Design. *Occupational Psychology*, 41(4), pp.211–221.

Apple Computer Inc., 1988. *Macintosh System Software 6*. 6 ed. [computer program]

Apple Inc., 2013. *OS X 10.9 Mavericks*. 10 ed. [computer program]

Apple Inc., 2016a. *Apple celebrates one billion iPhones*. Press Release, 27 July 2016

Apple Inc., 2016b. *iPhone User Guide for iOS 10.2*. [e-book]

Apple Inc., 2018. *Apple Reports Fourth Quarter Results*. Press Release, 01 November 2018

Apple Inc., 2019. *iPhone User Guide iOS 13.2*. Apple Inc. [e-book]

Bangor, A., Kortum, P.T. and Miller, J.T., 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), pp.574–594.

Bastien, J.M.C. and Scapin, D.L., 1995. Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*, 7(2), pp.105–121.

Bevan, N. and Macleod, M., 1994. Usability measurement in context. *Behaviour & Information Technology*, 13(1-2), pp.132–145.

Bevan, N., 1995. Measuring usability as quality of use. *Software Quality Journal*, 4(2), pp.115–130.

Bevan, N., Kirakowski, J. and Maissel, J., 1991. What is Usability? International Conference on HCI. pp.1–5.

Blackler, A., Popovic, V. and Mahar, D., 2003. The nature of intuitive use of products: an experimental approach. *Design Studies*, 24(6), pp.491–506.

Blackler, A., Popovic, V. and Mahar, D., 2010. Investigating users' intuitive interaction with complex artefacts. *Applied Ergonomics*, 41(1), pp.72–92.

Blandford, A. and Green, T., 2008. Methodological Development. In: P. Cairns and A.L. Cox, eds., *Research Methods for Human–Computer Interaction*. Cambridge: Cambridge University Press, pp.158–174.

Blandford, A.E., Hyde, J.K., Green, T.R.G. and Connell, I., 2008. Scoping analytical usability evaluation methods: a case study. *Human-Computer Interaction*, 23(3), pp.278–327.

Blanke, W. and Bajaj, C., 2002. Active Visualization in a Multidisplay Immersive Environment. In: S. Müller and W. Stürzlinger, eds. Eighth Eurographics Workshop on Virtual Environments. pp.1–10.

Blaschke, C.M., Freddolino, P.P. and Mullen, E.E., 2009. Ageing and Technology: A Review of the Research Literature. *British Journal of Social Work*, 39(4), pp.641–656.

Blessing, L., Chakrabarti, A. and Wallace, K.M., 1998. An overview of descriptive studies in relation to a general design research methodology. pp.42–56.

Blessing, L.T.M. and Chakrabarti, A., 2009. *DRM, a Design Research Methodology*. Springer Science & Business Media.

Blumer, H., 1969. *Symbolic interactionism: perspective and method*. Prentice-Hall.

Boehm, B.W. and Papaccio, P.N., 1988. Understanding and Controlling Software Costs. *IEEE Trans. Software Eng*, 14(10), pp.1462–1477.

Boehm, B.W., 2006. A View of 20th and 21st Century Software Engineering. *ICSE*, p.12.

Boote, D.N. and Beile, P., 2005. Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation. *Educational Researcher*, 34(6), pp.3–15.

Boote, D.N. and Beile, P., 2006. On 'Literature reviews of, and for, educational research': A response to the critique by Joseph Maxwell. *Educational Researcher*, 35(9), pp.32–35.

Booth, W.C., Colomb, G.G. and Williams, J.M., 2008. *The Craft of Research*. Third ed. University of Chicago Press.

Borsci, S., Macredie, R.D., Barnett, J., Martin, J., Kuljis, J. and Young, T., 2013. Reviewing and Extending the Five-User Assumption: A Grounded Procedure for Interaction Evaluation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(5), pp.29–23.

Box, G.E.P. and Draper, N.R., 1987. *Empirical model-building and response surfaces*. New York: John Wiley & Sons.

Box, G.E.P., 1976. Science and Statistics. *Journal of the American Statistical Association*, 71(356), pp.791–799.

Bradley, M., Hosking, I.M., Langdon, P.M. and Clarkson, P.J., 2017. Exploring Summative Depictions of Older User Experiences Learning and Adopting New Technologies. In: *Universal Access in Human–Computer Interaction. Design and Development Approaches and Methods*, 11th International Conference, UAHCI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part I. Cham: Springer, Cham, pp.21–30.

Bridgeman, B., Hendry, D. and Stark, L., 1975. Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, 15(6), pp.719–722.

Brocklehurst, P.R., Langley, J., Baker, S.R., McKenna, G., Smith, C. and Wassall, R., 2019. Promoting co-production in the generation and use of research evidence to improve service provision in special care dentistry. *British Dental Journal*, 227(1), pp.15–18.

Brooke, J., 1986. *System Usability Scale (SUS): A Quick-and-Dirty Method of System Evaluation User Information, Digital Equipment Co Ltd*. Reading.

Bush, V., 1945. As We May Think. *The Atlantic*. 1 Jul.

Campos, D.G., 2009. On the distinction between Peirce's abduction and Lipton's Inference to the best explanation. *Synthese*, 180(3), pp.419–442.

Caposella, C., 2005. PDC 2005 Keynote with Bill Gates. Professional Developers Conference. Los Angeles.

Capra, M.G. and Smith-Jackson, T.L., 2005. *Developing Guidelines for Describing Usability Problems*. Blacksburg, pp.1–43.

Capra, M.G., 2007. Comparing Usability Problem Identification and Description by Practitioners and Students. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 51(5), pp.474–477.

Card, S., Moran, T.P. and Newell, A., 1983a. The GOMS Model of Manuscript Editing. In: S.K. Card, T.P. Moran and A. Newell, eds., *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey.

Card, S.K., Moran, T.P. and Newell, A., 1983b. The Human Information-Processor. In: S.K. Card, T.P. Moran and A. Newell, eds., *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc.

Carroll, J.M., 2010. Conceptualizing a possible discipline of human-computer interaction. *Interacting with Computers*, 22(1), pp.3–12.

Carroll, J.M., Kellogg, W.A. and Rosson, M.B., 1991. The Task-Artifact Cycle. In: J.M. Carroll, ed., *Designing Interaction Psychology at the Human-Computer Interface*. Cambridge, pp.74–102.

CENELEC, 2015. *IEC 62366-1:2015 Part 1: Application of usability engineering to medical devices.* Brussels: European Committee for Electrotechnical Standardization

CENELEC, 2016. *IEC TR 62366-2 Part 2: Guidance on the application of usability engineering to medical devices.* Brussels: European Committee for Electrotechnical Standardization

Charmaz, K., 2011. Qualitative Interviewing and Grounded Theory Analysis. In: J.F. Gubrium and J.A. Holstein, eds., *Handbook of Interview Research*. 2455 Teller Road, Thousand Oaks California 91320 United States of America : SAGE Publications, Inc., pp.675–694.

Chathoth, P., Altinay, L., Harrington, R.J., Okumus, F. and Chan, E.S.W., 2013. Co-production versus co-creation: A process based continuum in the hotel service context. *International Journal of Hospitality Management*, 32, pp.11–20.

Chauhan, B.C., Tompkins, J.D., LeBlanc, R.P. and McCormick, T.A., 1993. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Investigative Ophthalmology & Visual Science*, 34(13), pp.3534–3540.

Chesterton, G.K., 1929. *The Thing*. 1st ed. London: Sheed & Ward.

Chin, J.P., Diehl, V.A. and Norman, K.L., 1988. *Development of an instrument measuring user satisfaction of the human-computer interface*. *the SIGCHI conference*, New York, New York, USA: ACM, pp.213–218.

Clarkson, P.J., Waller, S. and Cardoso, C., 2015. Approaches to estimating user exclusion. *Applied Ergonomics*, 46(PB), pp.304–310.

Constantine, L.L. and Lockwood, L.A.D., 1999. *Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design*. New York: Addison-Wesley.

Cooper, H.M., 1985. A Taxonomy of Literature Reviews. 69th Annual Meeting of the American Educational Research Association. Chicago, pp.1–49.

Coren, S. and Porac, C., 1974. The fading of stabilized images: Eye movements and information processing. *Perception & Psychophysics*, 16(3), pp.529–534.

Coryn, C., 2006. The Fundamental Characteristics of Research. *Journal of MultiDisciplinary Evaluation*, pp.124–133.

Costa, A.L. and Kallick, B., 2000. Discovering and Exploring Habits of Mind.

Creswell, J.W., 2009. *Research Design*. SAGE.

Crotty, M., 1998. *The Foundations of Social Research: Meaning and Perspective in the Research Process*. SAGE Publications.

Czaja, S.J., 2005. The impact of aging on access to technology. *ACM SIGACCESS Accessibility and Computing*, (83), pp.7–11.

Deane, J., Morland, R., Hosking, I. and Clarkson, J., 2011. *Building a Business Case for Inclusive Design*. Cambridge: University of Cambridge.

Debernardi, A., Sala, E., D'Aliberti, G., Talamonti, G., Franchini, A.F. and Collice, M., 2010. Alcmaeon of Croton. *Neurosurgery*, 66(2), pp.247–252.

Dedes, C., 2005. The Mechanism of Vision: Conceptual Similarities between Historical Models and Children's Representations. *Science & Education*, 14(7-8), pp.699–712.

DeWalt, K.M. and DeWalt, B.R., 2011. *Participant Observation: A Guide for Fieldworkers*. 2nd ed. AltaMira Press.

Dierckx de Casterlé, B., Gastmans, C., Bryon, E. and Denier, Y., 2011. QUAGOL: A guide for qualitative data analysis. *International Journal of Nursing Studies*, 49(3), pp.1–12.

Division, U.N.D.O.E.A.S.A.P., 2017. *World Population Prospects*. 2017 ed. New York: United Nations, Department of Economic and Social Affairs, Population Division.

Dix, A., Finlay, J., Abowd, G.D. and Beale, R., 2004. Socio-organizational issues and stakeholder requirements. In: *Human-Computer Interaction*, 3rd ed. London: Prentice Hall.

Dixon, J.R., 1966. *Design engineering: inventiveness, analysis and decision making.* New York: McGraw-Hill.

Dixon, J.R., 1987. On research methodology towards a scientific theory of engineering design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 1(3), pp.145–157.

Doll, W.J. and Torkzadeh, G., 1988. The Measurement of End-User Computing Satisfaction. *MIS Quarterly*, 12(2), pp.259–274.

Dorst, K. and Cross, N., 2001. Creativity in the design process: co-evolution of problem–solution. *Design Studies*, 22(5), pp.425–437.

Doug Engelbart Institute, 2018. Doug's Great Demo: 1968. [online] Doug Englebart Institute. Available at: <http://www.dougengelbart.org/firsts/dougs-1968-demo.html> [Accessed 28 Aug. 2018].

Dumas, J.S. and Redish, J., 1999. A Practical Guide to Usability Testing.

Eckstein, M.P., 2011. Visual search: A retrospective. *Journal of Vision*, 11(5), pp.14–14.

Egan, M.E. and Shera, J.H., 1952. Foundations of a Theory of Bibliography. *The Library Quarterly*, 22(2), pp.125–137.

Egeth, H.E., 1966. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1(4), pp.245–252.

Elliott, D., Whitaker, D. and MacVeigh, D., 1990. Neural contribution to spatiotemporal contrast sensitivity decline in healthy ageing eyes. *Vision Research*, 30(4), pp.541–547.

Emond, B. and West, R.L., 2004. Using Cognitive Modelling Simulations for User Interface Design Decisions. *Innovations in Applied Artificial Intelligence. IEAAIE . Lecture Notes in Computer Science*, 3029, pp.305–314.

English, W.K., Engelbart, D.C. and Berman, M.L., 1967. Display-Selection Techniques for Text Manipulation. *Human Factors in Electronics, IEEE Transactions on*, (1), pp.5–15.

Ethics Committee of the British Psychological Society, 2009. *Code of Ethics and Conduct*. Leicester, pp.1–29.

Farell, B. and Pelli, D.G., 1999. Psychophysical methods, or how to measure a threshold , and why. In: *Vision Research A Practical Guide to Laboratory Methods*. Oxford University Press, pp.129–136.

Faria, T.V.M., Pavanellie, M. and Bernardes, J.L., 2016. *Evaluating the Usability Using USE Questionnaire: Mindboard System Use Case*. Cham: Springer International Publishing, pp.1–10.

Faulkner, L., 2003. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments and Computers*, 35(3), pp.379–383.

Fernandez, A., Insfran, E. and Abrahão, S., 2011. Usability evaluation methods for the web: A systematic mapping study. *Information and Software Technology*, 53(8), pp.789–817.

Food and Drug Administration, 2016. *Applying Human Factors and Usability Engineering to Medical Devices.* Rockville: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Office of Device Evaluation

Fook, J., 2015. Reflective Practice and Critical Reflection. In: J. Lishman, ed., *Handbook for Practice Learning in Social Work and Social Care*, 3rd ed. books.google.com, pp.440–454.

Furman, S., Theofanos, M.F. and Wald, H., 2014. *Human Engineering Design Criteria Standards Part 1: Project Introduction and Existing Standards*. National Institute of Standards and Technology, pp.1–81.

Gentner, D. and Nielsen, J., 1996. The Anti-Mac Interface. *Communications of the ACM*, 39(8), pp.70–82.

Gero, J.S., 1990. Design Prototypes: A Knowledge Representation Schema for Design. *AI Magazine*, 11(4), p.26.

Gibson, J.J., 1966. *The senses considered as perceptual systems*. Boston, Houghton Mifflin.

Gibson, J.J., 1979. *The ecological approach to visual perception*. Boston : Houghton Mifflin.

Gillespie-Gallery, H., Konstantakopoulou, E., Harlow, J.A. and Barbur, J.L., 2013. Capturing Age-Related Changes in Functional Contrast Sensitivity With Decreasing Light Levels in Monocular and Binocular Vision. *Investigative Ophthalmology & Visual Science*, 54(9), pp.6093–24.

Glaser, B.G. and Strauss, A.L., 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. [online] Chicago: Aldine.

Glaser, B.G., 1978. *Theoretical sensitivity*. Mill Valley, Calif.: Sociology Press.

Glaser, B.G., 2007. Constructivist Grounded Theory? *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 3(3).

Glimcher, P.W., 2003. Eye Movements. In: *Fundamental Neuroscience*, 2nd ed. Elsevier Science, pp.873–893.

Goldman, A. and Blanchard, T., 2018. *The Stanford Encyclopedia of Philosophy*. 2018 ed. [online] Metaphysics Research Lab, Stanford University. Available at: <\url{https://plato.stanford.edu/archives/sum2018/entries/epistemology-social/>.

Google Inc., 2019a. *Google Maps App*. 5.31.10 ed.

Google Inc., 2019b. Google Maps Web. [online] Available at: <https://www.google.co.uk/maps> [Accessed 25 Nov. 2019].

GoPro Inc., 2015. *HERO4 Silver User Manual*. GoPro Inc., pp.1–94.

GoPro Inc., 2018. HERO4 Field of View (FOV) Information. [online] Available at: <https://gopro.com/help/articles/Question_Answer/HERO4-Field-of-View-FOV-Information> [Accessed 29 Dec. 2018].

Grant, M.J. and Booth, A., 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), pp.91–108.

Gray, W.D. and Salzman, M.C., 1998. Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods. *Human-Computer Interaction*, 13(3), pp.203–261.

Green, M., 2008. Physics, Physiology, Psychology. In: M. Green, J.M. Allen, S.B. Abrams and L. Weintraub, eds., *Forensic Vision with Application to Highway Safety*. Tucson: Lawyers & Judges Publishing Company, pp.49–81.

Green, M., Allen, J.M., Abrams, S.B. and Weintraub, L., 2008. *Forensic Vision with Application to Highway Safety*. 3rd ed. Tucson: Lawyers & Judges Publishing Company.

Green, T.R.G., 1989. Cognitive dimensions of notations. In: A. Sutcliffe and L. Macaulay, eds. People and computers V. pp.443–460.

Greene, J.C., 2006. Toward a methodology of mixed methods social inquiry. 13(1), pp.93–98.

Groarke, L., 2017. Informal Logic. In: E.N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, 2017 ed. Metaphysics Research Lab, Stanford University.

Gross, C.G., 1999. The fire that comes from the eye. *The Neuroscientist*, 5(1), pp.58–64.

Grudin, J., 1992. Utility and Usability - Research Issues and Development Contexts. *Interacting with Computers*, 4(2), pp.209–217.

Grudin, J., 2012. A moving target: The evolution of human-computer interaction. *The Human-Computer Interaction Handbook–Fundamentals, Evolving Technologies, and Emerging Applications*.

Grzybowski, A., 2009. Harry Moss Traquair (1875-1954), Scottish ophthalmologist and perimetrist. *Acta Ophthalmologica*, 87(4), pp.455–459.

Guba, E.G., 1990. *The Paradigm Dialog*. SAGE.

Gutwin, C., Cockburn, A. and Coveney, A., 2017. Peripheral Popout. the 2017 CHI Conference. New York, New York, USA: ACM Press, pp.208–219.

Ham, D.-H., 2013. A model-based framework for classifying and diagnosing usability problems. *Cognition, Technology & Work*, 16(3), pp.373–388.

Hansen, T., Pracejus, L. and Gegenfurtner, K.R., 2009. Color perception in the intermediate periphery of the visual field. *Journal of Vision*, 9(4), pp.26–26.

Harrison, R., Flood, D. and Duce, D., 2013. Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science*, 1(1), p.1.

Hart, S.G. and Staveland, L.E., 1988. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*. Advances in Psychology. *Advances in Psychology*, Elsevier, pp.139–183.

Hartson, H.R., Andre, T.S. and Williges, R.C., 2003. Criteria For Evaluating Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15(1), pp.145–181.

Hertzum, M. and Jacobsen, N.E., 2003a. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15(1), pp.183–204.

Hertzum, M. and Jacobsen, N.E., 2003b. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, 15(1), pp.183–204.

Hertzum, M., Molich, R. and Jacobsen, N.E., 2014. What you get is what you see: revisiting the evaluator effect in usability tests. *Behaviour & Information Technology*, 33(2), pp.144–162.

Hilbert, D.M. and Redmiles, D.F., 2000. Extracting usability information from user interface events. *ACM Computing Surveys*, 32(4), pp.384–421.

Hilbert, M. and López, P., 2011. The World's Technological Capacity to Store, Communicate, and Compute Information. *science*, 332(6025), pp.60–65.

Hinde, S. and Spackman, E., 2014. Bidirectional Citation Searching to Completion: An Exploration of Literature Searching Methods. *PharmacoEconomics*, 33(1), pp.5–11.

Hinrichs, S., n.d. *A Systems Approach to Improving Patient Safety through Medical Device Purchasing*. University of Cambridge, pp.1–346.

Hornbæk, K. and Frøkjær, E., 2008. A Study of the Evaluator Effect in Usability Testing. *Human-Computer Interaction*, 23(3), pp.251–277.

Hornbæk, K., 2006. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), pp.79–102.

Hosking, I., 2017. *iPhone Visibility Analysis R1*.

Hosking, I., Waller, S. and Clarkson, P.J., 2010. It is normal to be different: Applying inclusive design in industry. *Interacting with Computers*, 22(6), pp.496–501.

Hosking, I.M. and Clarkson, P.J., 2017. Now You See It, Now You Don't: Understanding User Interface Visibility. In: M. Winckler, H. Johnson and P. Palanque, eds., *Task Models and Diagrams for User Interface Design*, Lecture Notes in Computer Science. Cham: Springer International Publishing, pp.436–445.

Hosking, I.M. and Clarkson, P.J., 2018a. Believing is Seeing: Rethinking the Visibility of User Interfaces. In: P.M. Langdon, J. Lazar, A. Heylighen and H. Dong, eds. Cambridge Workshop on Universal Access and Assistive Technology. Cambridge, pp.3–13.

Hosking, I.M. and Clarkson, P.J., 2018b. vis-UI-lise: developing a tool for assessing user interface visibility. In: T.Z. Ahram and C. Falcão, eds. AHFE 2018 International Conferences on Usability User Experience and Human Factors and Assistive Technology. The 9th International Conference on Applied Human Factors and Ergonomics.

Högberg, D., Brolin, E. and Hanson, L., 2015. Identification of redundant boundary cases. *Proceedings of the 19th Triennial Congress of the International Ergonomics Association*.

Hurtienne, J., Horn, A.-M., Langdon, P.M. and Clarkson, P.J., 2013. Facets of prior experience and the effectiveness of inclusive design. *Universal Access in the Information Society*, 12(3), pp.297–308.

Hutchins, E.L., Hollan, J.D. and Norman, D.A., 1985. Direct Manipulation Interfaces. *Human-Computer Interaction*, 1(4), pp.311–338.

International Organization for Standardization, 2010. *EN ISO 9241-100:2010 . Ergonomics of human-system interaction - Part 100: Introduction to standards related to software ergonomics*. Brussels: European Committee for Standardization.

International Organization for Standardization, 2011. *EN ISO 9241-303:2011Ergonomics of human-system interaction - Part 303: Requirements for electronic visual displays*. Brussels: European Committee for Standardization, pp.1–43.

International Organization for Standardization, 2016. *EN ISO 9241-161:2016 Ergonomics of human-system interaction - Part 161: Guidance on visual user-interface elements*. Brussels: European Committee for Standardization, pp.1–69.

International Organization for Standardization, 2017a. *EN ISO 9241-112:2017 Ergonomics of human-system interaction - Part 112: Principles for the presentation of information*. Brussels: European Committee for Standardization, pp.1–30.

International Organization for Standardization, 2017b. *EN ISO 9241-125:2017 Ergonomics of human-system interaction Part 125: Guidance on visual presentation of information*. Brussels: European Committee for Standardization, pp.1–40.

International Organization for Standardization, 2018. *EN ISO 9241-306:2018 Ergonomics of human-system interaction - Part 306: Field assessment methods for electronic visual displays*. Brussels: European Committee for Standardization, pp.1–54.

Irby, C., Bergsteinsson, L., Moran, T.P., Newman, W. and Tesler, L., 1977. *A Methodology for User Interface Design*. Xerox Corporation, pp.1–54.

Ivory, M.Y. and Hearst, M.A., 2001. The State of the Art in Automating Usability Evaluation of User Interfaces. *ACM Computing Surveys*, 33(4), pp.470–516.

Jacobsen, N.E., Hertzum, M. and John, B.E., 1998. The Evaluator Effect in Usability Studies: Problem Detection and Severity Judgments. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 42(19), pp.1336–1340.

James, W., 1890. *The Principles of Psychology*. NY, US: Henry Holt and Company.

John, B.E. and Kieras, D.E., 1996. The GOMS Family of User Interface Analysis Techniques - Comparison and Contrast. *ACM Transactions on Computer-Human Interaction*, 3(4), pp.320–351.

Johnson, R.H., 2014. The Rise of Informal Logic. *Windsor Studies in Argumentation*, 2, pp.1–269.

Karat, C.-M., 1990. Cost-Benefit Analysis of Usability Engineering Techniques. *Proceedings of the Human Factors Society Annual Meeting*, 34(12), pp.839–843.

Karat, J., 1997. Chapter 28 - User-Centered Software Evaluation Methodologies. In: M.G. Helander, T.K. Landauer and P.V. Prabhu, eds., *Handbook of Human-Computer Interaction (Second Edition)*, Handbook of Human-Computer Interaction (Second Edition). Amsterdam: North-Holland, pp.689–704.

Karray, F., Alemzadeh, M. and Saleh, J.A., 2008. Human-Computer Interaction: Overview on State of the Art. *International Journal on Smart Sensing and Intelligent Systems*, 1(1), pp.137–159.

Katsanos, C., Tselios, N.K. and Avouris, N.M., 2010. A Survey of Tools Supporting Design and Evaluation of Websites Based on Models of Human Information Interaction. *International Journal on Artificial Intelligence Tools*, 19(06), pp.755–781.

Kay, A.C., 1972. A Personal Computer for Children of All Ages. Proceedings of the ACM annual conference. New York, NY, USA.

Keates, S.L. and Clarkson, P.J., 2004. *Countering Design Exclusion*. Springer-Verlag.

Keenan, S.L., Hartson, H.R., Kafura, D.G. and Schulman, R.S., 1999. The Usability Problem Taxonomy: A Framework for Classification and Analysis. *Empirical Software Engineering*, 4(1), pp.71–104.

Kirakowski, J. and Corbett, M., 1993. SUMI: the Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3), pp.210–212.

Kolko, J., 2010. Abductive Thinking and Sensemaking: The Drivers of Design Synthesis. *Design Issues*, 26(1), pp.15–28.

Koretz, J.F., Kaufman, P.L., Neider, M.W. and Goeckner, P.A., 1989. Accommodation and presbyopia in the human eye - aging of the anterior segment. *Vision Research*, 29(12), pp.1685–1692.

Koutsabasis, P., Spyrou, T. and Darzentas, J., 2007. Evaluating Usability Evaluation Methods: Criteria, Method and a Case Study. In: J.A. Jacko, ed. Berlin, Heidelberg: Springer Berlin Heidelberg, pp.569–578.

Kuhn, T.S., 1962. *The Structure of Scientific Revolutions*. 1st ed. Chicago: University of Chicago Press.

Kurzweil, R., 2005. *The singularity is near : when humans transcend biology*. New York: Viking.

LaBerge, D., 1983. Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance*, 9(3), pp.371–379.

Landauer, T.K., 1995. *The Trouble with Computers*. MIT Press.

Langdon, P., Lewis, T. and Clarkson, J., 2007. The effects of prior experience on the use of consumer products. *Universal Access in the Information Society*, 6(2), pp.179–191.

Lave, J. and Wenger, E., 1991. Situated Learning: Legitimate Peripheral Participation.

Lavery, D., Cockton, G. and Atkinson, M.P., 1997. Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, 16(4-5), pp.246–266.

Lawton, M.P. and Brody, E.M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. *The Gerontologist*, 9(3), pp.179–186.

Legge, G.E., Rubin, G.S. and Luebker, A., 1987. Psychophysics of Reading-V. The of Contrast in Normal Vision. *Vision Research*, 27(7), pp.1165–1177.

Lewis, J.R., 1991. An After-Scenario Questionnaire for Studies: Psychometric Evaluation Over Three Trials. *ACM SIGCHI Bulletin*, 23(4), p.79.

Lewis, J.R., 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), pp.57–78.

Licklider, J.C.R., 1960. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1(1), pp.4–11.

Liu, Z., 2004. Perceptions of credibility of scholarly information on the web. *Information Processing & Management*, 40(6), pp.1027–1038.

Lou Maher, M., Poon, J. and Boulanger, S., 1996. Formalising Design Exploration as Co-Evolution. In: *Advances in Formal Design Methods for CAD*. Boston, MA: Springer, Boston, MA, pp.3–30.

Lund, A.M., 2001. Measuring usability with the USE questionnaire. *Usability Interface*, 8(2), pp.3–6.

Macleod, M. and Rengger, R., 1993. *The Development of DRUM: A Software Tool for Video-assisted Usability Evaluation*. *People and Computers VIII*, Cambridge University Press, pp.293–309.

Manakhov, P. and Ivanov, V.D., 2016. Defining Usability Problems. the 2016 CHI Conference Extended Abstracts. New York, New York, USA: ACM Press, pp.3144–3151.

Mantei, M.M. and Teorey, T.J., 1988. Cost/benefit analysis for incorporating human factors in the software lifecycle. *Communications of the ACM*, 31(4), pp.428–439.

Marr, D. and Nishihara, H.K., 1978. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B, Biological sciences*, 200(1140), pp.269–294.

Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.

Masuda, T. 2010. 'Cultural Effects on Visual Perception', in *Encyclopaedia of Perception*. Sage, London

McMains, S.A. and Somers, D.C., 2004. Multiple spotlights of attentional selection in human visual cortex. *Neuron*, 42(4), pp.677–686.

Mejdal, S. and McCauley, M.E., 2001. *Human Factors Design Guidelines for Multifunction Displays*. pp.1–79.

Menzies, P., 2017. Counterfactual Theories of Causation. [online] Available at: <https://plato.stanford.edu/archives/win2017/entries/causation-counterfactual/>.

Merriam, S.B., 2002. Assessing and Evaluating Qualitative Research. In: S.B. Merriam, ed., *Qualitative Research in Practice*. Jossey-Bass.

Microsoft, 2014. *User Guide Nokia Lumia 830*. Microsoft, pp.1–118.

Mukaka, M.M., 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal : the journal of Medical Association of Malawi*, 24(3), pp.69–71.

Myers, B.A., 1998. A brief history of human-computer interaction technology. *ACM Interactions*, 5(2), pp.44–54.

Neisser, U., 1976. *Cognition and Reality: Principles and Implications of Cognitive Psychology*. W H Freeman & Company.

Newell, A., Gregor, P., Morgan, M., Pullin, G. and Macaulay, C., 2010. User-Sensitive Inclusive Design. *Universal Access in the Information Society*, 10(3), pp.235–243.

Nicholl, B., 2017. Empathy as an Aspect of Critical Thought and Action in Design and Technology. In: *Critique in Design and Technology Education*, Contemporary Issues in Technology Education, 4 ed. Singapore: Springer, Singapore, pp.153–171.

Nielsen, J. and Landauer, T.K., 1993. A mathematical model of the finding of usability problems. *INTERCHI*, pp.206–213.

Nielsen, J. and Mack, R.L., 1994. *Usability Inspection Methods*. Wiley.

Nielsen, J., 1993a. Noncommand user interfaces. *Communications of the ACM*, 36(4), pp.83–99.

Nielsen, J., 1993b. *Usability Engineering*. San Diego: Morgan Kaufmann.

Nielsen, J., 1994. *Usability Engineering*. Morgan Kaufmann.

Noelker, L.S. and Browdie, R., 2014. Sidney Katz, MD: A New Paradigm for Chronic Illness and Long-Term Care. *The Gerontologist*, 54(1), pp.13–20.

Norman, D.A. and Nielsen, J., 2010. Gestural interfaces: a step backward in usability. *ACM Interactions*, 17(5), pp.46–49.

Norman, D.A., 1986. Cognitive Engineering. In: D.A. Norman and S.W. Draper, eds., *User Centered System Design New Perspectives on Human-Computer Interaction*. Hillsdale, NJ, USA, pp.31–62.

Norman, D.A., 1988. *The Psychology of Everyday Things*. New York, NY: Basic Books.

Norman, D.A., 1999. Affordance, conventions, and design. *ACM Interactions*, 6(3), pp.38–43.

Norman, D.A., 2002. The Design of Everyday Things. pp.1–270.

Norman, D.A., 2010. Natural user interfaces are not natural. *ACM Interactions*, 17(3), pp.6–10.

Norman, D.A., 2011. Living with Complexity.

Norman, D.A., 2013. *The Design of Everyday Things*. 3rd ed. New York: Basic Books.

Norman, K.L., Shneiderman, B., Harper, B. and Slaughter, L., 2007. *Questionnaire for User Interface Satisfaction*. 7 ed.

North, J.D., 1954. *The rational behavior of mechanically extended man*. Wolverhampton: Boulton Paul Aircraft Ltd.

Nyamsuren, E. and Taatgen, N.A., 2013. Pre-attentive and attentive vision module. *Cognitive Systems Research:Special Issue on ICCM2012*, 24, pp.62–71.

OECD, 2002. *Frascati Manual*. 6 ed. [online] Organisation for Economic Co-operation and Development. Available at: <http://www.oecd.org/innovation/inno/frascatimanualproposedstandardpracticeforsurveysonresearchandexperimentaldevelopment6thedition.htm>.

Ohnemus, K.R. and Biers, D.W., 1993. Retrospective versus Concurrent Thinking-Out-Loud in Usability Testing. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 37(17), pp.1127–1131.

Ormerod, T.C., Richardson, J. and Shepherd, A., 2000. Enhancing the usability of a task analysis method: a notation and environment for requirements specification. In: J. Annett and N. Stanton, eds., *Task Analysis*. London and New York: Taylor & Francis, pp.114–135.

Pahl, G. and Beitz, W., 1988. *Engineering design : a systematic approach.* A. Pomernans and K.M. Wallace London: Design Council.

Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N. and Hoagwood, K., 2015. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), pp.533–544.

Pankok, C. and Kaber, D.B., 2018. An integrated measure of display clutter based on feature content, user knowledge and attention allocation factors. *Ergonomics*, 61(5), pp.682–696.

Patton, L., 2016. Hermann von Helmholtz. [online] plato.stanford.edu. Available at: <<https://plato.stanford.edu/archives/win2016/entries/hermann-helmholtz/>.> [Accessed 10 Aug. 2017].

Pawson, R., Greenhalgh, T., Harvey, G. and Walshe, K., 2005. Realist review--a new method of systematic review designed for complex policy interventions. *Journal of health services research & policy*, 10 Suppl 1(1_suppl), pp.21–34.

Penny, R.K., 1970. Principles of Engineering Design. *Postgraduate Medical Journal*, 46(536), pp.344–&.

Perez, C., 2010. Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, 34(1), pp.185–202.

Persad, U., Langdon, P. and Clarkson, J., 2007. Characterising user capabilities to support inclusive design evaluation. *Universal Access in the Information Society*, 6(2), pp.119–135.

National Information Board, 2014. *Personalised Health and Care 2020*. HM Government, pp.1–66.

Persson, H., Åhman, H., Yngling, A.A. and Gulliksen, J., 2015. Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. *Universal Access in the Information Society*, 14(4), pp.1–22.

Peute, L.W.P., de Keizer, N.F. and Jaspers, M.W.M., 2015. The value of Retrospective and Concurrent Think Aloud in formative usability testing of a physician data query tool. *Journal of Biomedical Informatics*, 55(C), pp.1–10.

Pinto, M.R., De Medici, S., Zlotnicki, A., Bianchi, A., Van Sant, C. and Napoli, C., 1997. Reduced visual acuity in elderly people: the role of ergonomics and gerontechnology. *Age and ageing*, 26(5), pp.339–344.

Posner, M.I., Snyder, C.R. and Davidson, B.J., 1980. Attention and the detection of signals. *Journal of experimental psychology*, 109(2), pp.160–174.

Rajeshkumar, S., Omar, R. and Mahmud, M., 2013. Taxonomies of User Experience (UX) evaluation methods. 2013 International Conference on Research and Innovation in Information Systems (ICRIIS). IEEE, pp.533–538.

Redish, J. and Wixon, D., 2002. Task analysis. In: J.A. Jacko and A. Sears, eds., *The Human-Computer Interaction Handbook*. Mahwah, New Jersey: CRC PressI Llc, pp.922–940.

Rieger, K.L., 2018. Discriminating among grounded theory approaches. *Nursing Inquiry*, 26(1), pp.e12261–12.

Rieh, S.Y., 2002. Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, 53(2), pp.145–161.

Ritchie, H. and Roser, M., 2017. Technology Adoption. *Our World in Data*. [online] Available at: <https://ourworldindata.org/technology-adoption>.

Roberson, D. and Hanley, J.R., 2007. Color Vision: Color Categories Vary with Language after All. *Current Biology*, 17(15), pp.R605–R607.

Roberson, D., Davidoff, J., Davies, I.R.L. and Shapiro, L.R., 2005. Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50(4), pp.378–411.

Rocco, T.S. and Plakhotnik, M.S., 2009. Literature Reviews, Conceptual Frameworks, and Theoretical Frameworks: Terms, Functions, and Distinctions. *Human Resource Development Review*, 8(1), pp.120–130.

Rogers, Y., 2004. New theoretical approaches for human-computer interaction. *Annual Review of Information Science and Technology*, 38(1), pp.87–143.

Rogers, Y., Sharp, H. and Preece, J., 2011. *Interaction design: beyond human-computer interaction ; 3rd ed. Online Information Review*, Chichester West Sussex UK: J Wiley & Sons, pp.102–114.

Rosenholtz, R., Li, Y., Mansfield, J. and Jin, Z., 2005. Feature congestion: a measure of display clutter. *CHI*, p.761.

Rubin, J. and Chisnell, D., 2008. *Handbook of Usability Testing : How to Plan, Design, and Conduct Effective Tests*. 2nd ed. [online] Indianapolis, IN : Wiley Pub. Available at: <https://books.google.co.uk/books/about/Handbook_of_Usability_Testing.html?id=l_e1MmVzMb0C&redir_esc=y>.

Runco, M. and Scott, K.P., 2014. *Human Integration Design Handbook (HIDH)*. 1st ed. Washington, DC, pp.1–1294.

Russell, C.L., 2005. An overview of the integrative research review. *Progress in transplantation (Aliso Viejo, Calif.)*, 15(1), pp.8–13.

Saraee, E., Jalal, M. and Betke, M., 2018. *SAVOIAS: A Diverse, Multi-Category Visual Complexity Dataset*. *arXiv.org*.

Sauro, J., 2014. The relationship between problem frequency and problem severity in usability evaluations. *Journal of Usability Studies*, 10(1), pp.17–25.

Schaffer, J., 2016. The Metaphysics of Causation. [online] plato.stanford.edu. Available at: <https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/> [Accessed 3 Jul. 2019].

Schmettow, M., 2008. Heterogeneity in the Usability Evaluation Process. *Proceedings of the 22nd British HCI Group Annual Conference on HCI 2008: People and Computers XXII: Culture, Creativity, Interaction*, 1, pp.89–98.

Schryen, G., Wagner, G. and Benlian, A., 2015. Theory of Knowledge for Literature Reviews: An Epistemological Model, Taxonomy and Empirical Analysis of IS Literature. Thirty Sixth International Conference on Informations Systems. Fort Worth, pp.1–22.

Sears, A., 1997. Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human-Computer Interaction*, 9(3), pp.213–234.

Seffah, A., Donyaee, M., Kline, R.B. and Padda, H.K., 2006. Usability measurement and metrics: A consolidated model. *Software Quality Journal*, 14(2), pp.159–178.

Sekuler, R. and Sekuler, A.B., 2000. Visual perception and cognition. In: L.K. Bowker, J.D. Price and S.C. Smith, eds., *Oxford textbook of Geriatric Medicine*, 2nd ed. pp.874–880.

Shackel, B., 1991. Usability - Context, framework, definition, design and evaluation. In: B. Shackel and S.J. Richardson, eds., *Human Factors for Informatics Usability*. New York, NY, USA: Cambridge University Press, pp.21–37.

Shepherd, A., 1998. HTA as a framework for task analysis. *Ergonomics*, 41(11), pp.1537–1552.

Shneiderman, B. and Bederson, B.B., 2003. *The Craft of Information Visualization: Readings and Reflections*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Shneiderman, B. and Plaisant, C., 2005. *Designing the user interface - strategies for effective human-computer interaction, 4th Edition.*

Shneiderman, B., 1983a. Direct manipulation: A step beyond programming languages. *Computer*, 16(8), pp.57–69.

Shneiderman, B., 1983b. Human factors of interactive software. In: *Enduser Systems and Their Human Factors*, Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, Berlin, Heidelberg, pp.9–29.

Shneiderman, B., 2000. Universal Usability: Pushing Human-Computer Interaction Research to Empower Every Citizen. *Communications of the ACM*, [online] 43(5), pp.84–91. Available at: <http://dl.acm.org/citation.cfm?doid=332833.332843>.

Shneiderman, B., 2003. Promoting universal usability with multi-layer interface design. *SIGCAPH Comput. Phys. Handicap*, (73-74), pp.1–8.

Simons, D.J. and Chabris, C.F., 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, 28(9), pp.1059–1074.

Smith, D.C., Irby, C., Kimball, R., Verblank, B. and Harslem, E., 1982. Designing the Star User Interface. *Byte*, (4), 1 Apr., pp.242–282.

Solso, R.L., 1996. *Cognition and the Visual Arts*. MIT Press.

Staplin, L., Lococo, K., Sim, J. and Drapcho, M., 1989. *Age differences in a visual information processing capability underlying traffic control device usage*. *Transportation Research Information Database*, pp.63–72.

Stember, M., 1991. Advancing the social sciences through the interdisciplinary enterprise. *The Social Science Journal*, 28(1), pp.1–14.

Stigmar, H. and Harrie, L., 2013. Evaluation of Analytical Measures of Map Legibility. *The Cartographic Journal*, 48(1), pp.41–53.

Stoller, K., 2018. Global 2000: The World's Largest Public Companies 2018.

Strasburger, H., Rentschler, I. and Juttner, M., 2011. Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5), pp.13–13.

Sukha, A.Y. and Rubin, A., 2013. Psychophysical aspects of contrast sensitivity. *African Vision and Eye Health*, 72(2), pp.76–85.

Sutherland, I.E., 1963. *Sketchpad, A Man-Machine Graphical Communication System*. pp.1–177.

Toulmin, S.E., 1958. *The Uses of Argument*. Cambridge University Press.

Toyota, 2010. *Avensis Owner's Manual*. 20 ed. Japan: Toyota Motor Corporation.

Traquair, H.M., 1927. *An introduction to clinical perimetry*. 1st ed. London: Henry Kimpton.

Treisman, A.M. and Gelade, G., 1980. A Feature-Integration Theory of Attention. *Cognitive Psychology*, 12(1), pp.97–136.

Tseng, S. and Fogg, B.J., 1999. Credibility and computing technology - Users want to trust, and generally do. But that trust is undermined, often forever, when the system delivers erroneous information. *Communications of the ACM*, 42(5), pp.39–44.

Tsoucalas, G., Karamanou, M., Kousoulis, A.A. and Androutsos, G., 2012. Democritus (460-370 BC) on embryology, anatomy and pediatrics: the unknown aspects of the Greek atomic scientist. *Italian journal of anatomy and embryology,*, 117(3), pp.199–204.

Turner, C., Lewis, J. and Nielsen, J., 2006. Determining Usability Test Sample Size. *International Encyclopedia of Ergonomics and Human Factors, Second Edition - 3 Volume Set*, 3, pp.3084–3088.

UK Digital Strategy.HM Government.

United Nations Department of Economic Affairs Population Division, 2009. *World Population Ageing 2009*. New York: United Nations, p.82.

University of Cambridge, 2019. Research Operations Office: Frascati Definition of Research. [online] admin.cam.ac.uk. Available at: <http://www.admin.cam.ac.uk/offices/research/Research/Frascati.aspx> [Accessed 30 Sep. 2019].

US Department of Defense, 2012. *Department of Defense Design Criteria Standard: Human Engineering*. p.382.

van Dam, A., 1997. Post-WIMP user interfaces. *Communications of the ACM*, 40(2), pp.63–67.

van Dam, A., 2001. User interfaces: disappearing, dissolving, and evolving. 44(3), pp.50–52.

van den Berg, R., Cornelissen, F.W. and Roerdink, J.B.T.M., 2009. A crowding model of visual clutter. *Journal of Vision*, 9(4), pp.24–24.

Van den Haak, M.J. and De Jong, M.D.T., 2003. *Exploring two methods of usability testing: concurrent versus retrospective think-aloud protocols*. IEEE.

van Veenendaal, E.P.W.M., 1998. Questionnaire based usability testing. European Software Quality Week. pp.1–9.

Vermeeren, A.P.O.S., van Kesteren, I. and Bekker, M.M., 2003. Managing the 'Evaluator Effect' in User Testing. *INTERACT*.

Virzi, R.A., 1992. Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(4), pp.457–468.

Voorberg, W.H., Bekkers, V.J.J.M. and Tummers, L.G., 2014. A Systematic Review of Co-Creation and Co-Production: Embarking on the social innovation journey. *Public Management Review*, 17(9), pp.1333–1357.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J.R., van der Helm, P.A. and van Leeuwen, C., 2012. A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, 138(6), pp.1218–1252.

Waller, S., Bradley, M., Hosking, I. and Clarkson, P.J., 2013. Making the case for inclusive design. *Applied Ergonomics*, pp.1–7.

Waller, S., Langdon, P. and Clarkson, P.J., 2009. Visualizing Design Exclusion Predicted by Disability Data: A Mobile Phone Case Study. In: C. Stephanidis, ed. Universal Access in Human-Computer Interaction 2009. Berlin, Heidelberg: Springer Berlin Heidelberg, pp.644–653.

Ware, C., 2003. Design as Applied Perception. In: J.M. Carroll, ed., *HCI models, theories, and frameworks : toward a multidisciplinary science*. San Francisco: Morgan Kaufmann, pp.11–26.

Ware, C., 2004. *Information Visualization : Perception for Design*. 2nd ed. San Francisco : Morgan Kaufman.

Watson, A.B. and Ahumanda, A.J., 2016. The pyramid of visibility. Human Vision and Electronic Imaging.

Watson, A.B., 2009. The Windows of Visibility. [online] NASA Human System Integraton Division. Available at: <https://human-factors.arc.nasa.gov/techareas/tech_faq/watson_visibility.php> [Accessed 9 Oct. 2018].

Watson, A.B., Ahumada, A.J. and Farrell, J.E., 1986. Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A: Optics*, 3(3), pp.300–307.

Wenger, E., 1999. *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

Whittemore, R. and Knafl, K., 2005. The integrative review: updated methodology. *Journal of Advanced Nursing*, 52(5), pp.546–553.

Wickens, C.D., Hollands, J.G., Banbury, S. and Parasuraman, R., 2013. *Engineering Psychology and Human Performance*. Pearson Education.

Wilson, P., 1983. *Second-hand knowledge : an inquiry into cognitive authority*. Westport, Conn. : Greenwood Press.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R. and Boroditsky, L., 2007. Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), pp.7780–7785.

Winer, G.A., Cottrell, J.E., Gregg, V., Fournier, J.S. and Bica, L.A., 2002. Fundamentally misunderstanding visual perception: Adults' belief in visual emissions. *American Psychologist*, 57(6-7), pp.417–424.

Wixon, D. and Wilson, C., 1997. Chapter 27 - The Usability Engineering Framework for Product Design and Evaluation. In: M.G. Helander, T.K. Landauer and P.V. Prabhu, eds., *Handbook of Human-Computer*

Interaction (Second Edition), Handbook of Human-Computer Interaction (Second Edition). Amsterdam: North-Holland, pp.653–688.

Wolfe, J., Cain, M., Ehinger, K. and Drew, T., 2015. Guided Search 5.0: Meeting the challenge of hybrid search and multiple-target foraging. *Journal of Vision*, 15(12), p.1106.

Wolfe, J.M. and Horowitz, T.S., 2004. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6), pp.495–501.

Wolfe, J.M. and Horowitz, T.S., 2017. Five factors that guide attention in visual search. *Nature Publishing Group*, 1(3), pp.1–8.

Wolfe, J.M., 1998. What Can 1 Million Trials Tell Us About Visual Search? *Psychological Science*, 9(1), pp.33–39.

Wolfe, J.M., 2007. Guided Search 4.0. In: W.D. Gray, ed., *Integrated Models of Cognitive Systems*, Current Progress With a Model of Visual Search. Oxford University Press.

Wolffsohn, J.S. and Davies, L.N., 2018. Presbyopia: Effectiveness of correction strategies. *Progress in Retinal and Eye Research*, 68, pp.1–20.

Wood, N.V., 2004. *Perspectives on Argument*. 4 ed. New Jersey: Pearson Education Inc.

Wright, P. and McCarthy, J., 2010. *Experience-Centered Design: Designers, Users, and Communities in Dialogue*. Morgan & Claypool

APPENDIX A

Evaluation of Windows Phone 10 Scenario using vis-UI-lise

A full copy of the evaluation of a Windows Phone 10 using the vis-UI-lise tool can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft PowerPoint but is stored there as a PDF to ensure readability across different platforms and over time.

Reference details:

Repository Title: vis-UI-lise Evaluation Tool Documentation R1

DOI link: https://doi.org/10.17863/CAM.48850

File name: A. Evaluation of Windows Phone 10 Scenario using vis-UI-lise

APPENDIX B

vis-UI-lise Training Presentation

A full copy of the vis-UI-lise Training Presentation can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft PowerPoint but is stored there as a PDF to ensure readability across different platforms and over time. Various images were substituted with near equivalent images to ensure there were appropriate image rights for this publication.

Reference details:

Repository Title: vis-UI-lise Evaluation Tool Documentation R1

DOI link: https://doi.org/10.17863/CAM.48850

File name: B. vis-UI-lise Training Presentation

APPENDIX C

vis-UI-lise Evaluator Guide

A full copy of the vis-UI-lise Evaluator Guide can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft PowerPoint but is stored there as a PDF to ensure readability across different platforms and over time.

Reference details:

Repository Title: vis-UI-lise Evaluation Tool Documentation R1

DOI link: https://doi.org/10.17863/CAM.48850

File name: C. vis-UI-lise Evaluator Guide

APPENDIX D

vis-UI-lise Quick Start Guide

A full copy of the vis-UI-lise Quick Start Guide can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft PowerPoint but is stored there as a PDF to ensure readability across different platforms and over time.

Reference details:

Repository Title: vis-UI-lise Evaluation Tool Documentation R1

DOI link: https://doi.org/10.17863/CAM.48850

File name: D. vis-UI-lise Quick Start Guide

APPENDIX E

vis-UI-lise Evaluation Template

A full copy of the vis-UI-lise Evaluation Template can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft PowerPoint but is stored there as a PDF to ensure readability across different platforms and over time. In practice a template like this is intended to be distributed and used in its source form.

Reference details:

Repository Title: vis-UI-lise Evaluation Tool Documentation R1

DOI link: https://doi.org/10.17863/CAM.48850

File name: E. vis-UI-lise Evaluation Template

APPENDIX F

vis-UI-lise Practice Exercise

A full copy of the vis-UI-lise Practice Exercise can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft PowerPoint but is stored there as a PDF to ensure readability across different platforms and over time. In practice a document like this is intended to be distributed and used in its source form.

Reference details:

Repository Title: vis-UI-lise Evaluation Tool Documentation R1

DOI link: https://doi.org/10.17863/CAM.48850

File name: F. vis-UI-lise Practice Exercise

APPENDIX G

Selected Interview Responses Regarding Key vis-UI-lise Terms

A full copy of the selected interview responses regarding key vis-UI-lise terms can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft Word but is stored there as a PDF to ensure readability across different platforms and over time.

Reference details:

Repository Title: vis-UI-lise PhD Supporting Documents

DOI link: https://doi.org/10.17863/CAM.48981

File name: Interview Responses Regarding Key vis-UI-lise Terms R1

APPENDIX H

Modified Technology Familiarity Questionnaire

A full copy of the modified Technology Familiarity Questionnaire can be found on the University of Cambridge's institutional repository called Apollo. The original document was authored in Microsoft Word but is stored there as a PDF to ensure readability across different platforms and over time.

Reference details:

Repository Title: vis-UI-lise PhD Supporting Documents

DOI link: https://doi.org/10.17863/CAM.48981

File name: Interview Responses Regarding Key vis-UI-lise Terms R1