The Formation of Objects: Reflecting on Human Factors and Safety Science from a Foucauldian Perspective

Author

Van Winsen, Roeland Dankmar

Published

2015

Thesis Type

Thesis (PhD Doctorate)

School

School of Humanities

DOI

10.25904/1912/1043

The formation of objects: Reflecting on human factors and safety science from a Foucauldian perspective

Roeland Dankmar van Winsen

MSc. Cognitive Psychology

MSc. Human Factors and Systems Safety

School of Humanities

Arts, Education and Law

Griffith University

Submitted in fulfilment of the requirements of the degree of

Doctor of Philosophy

February 2015

‘What, do you imagine that I would take so much trouble and so much pleasure in writing, do you think that I would keep so persistently to my task, if I were not preparing – with a rather shaky hand – a labyrinth into which I can venture, in which I can move my discourse, opening up underground passages, forcing it to go far from itself, finding overhangs that reduce and deform its itinerary, in which I can lose myself and appear at last to eyes that I will never have to meet again. I am no doubt not the only one who writes in order to have no face. Do not ask who I am and do not ask me to remain the same: leave it to our bureaucrats and our police to see that our papers are in order. At least spare us their morality when we write.’

Michel Foucault, The Archaeology of Knowledge, 2002, p. 19

Abstract

This thesis explores the possibility of drawing on a Foucauldian archaeological approach to critically reflect on some of the discursive practices that the disciplines of human factors and safety science are promoting. As such, it asks what insights Foucault's archaeological analysis of objects can provide these practically oriented disciplines. Specifically, this thesis develops and presents a number of ways to apply Foucault's archaeological level of description to study the constitution and discursive functioning of some prominent objects of the human factors and safety sciences discourses: it chiefly focuses on situation awareness (SA) and later expands the scope by analysing safety culture and resilience as objects of knowledge.

Through their archaeological description of the constitution of situation awareness, in terms of the 'rules of formation' and 'condition of possibilities' for the formation of the object, the first two studies argue that situation awareness is not a natural object, but rather a historically contingent object that emerges as both a condition and an effect of the need to make the human part of our socio-technological systems more susceptible to analysis. Similarly, the studies into safety culture and resilience describe the constitution of their respective objects and then go on to discuss the effects that these objects have. Both studies conclude that the objects function in a manner to place the responsibility for safety on individual workers.

One conclusion to draw from the archaeological analyses of the objects of situation awareness, safety culture, and resilience, is that these objects discursively turn the focus (of safety) onto the human individual as the locus for intervention, despite their explicit commitment to a systems approach. These are the kind of 'insights' that Foucault's archaeological analysis can provide the disciplines of human factors.

In aiming to build bridges between Foucault's archaeological studies and the disciplines of human factors and safety science, I hope that this thesis provides a small step in showing that Foucauldian (archaeological) analysis can create 'insights' and supply new ways of thinking that can contribute to the epistemological pluralism of human factors and safety science.

Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Roel van Winsen

February 2015

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisory team: Professor Sidney Dekker, Dr. Susanna Chamberlain, and Dr. Hugh Breakey. Sidney, after many years I have much to thank you for. Most prominently, without you I would not have had the pleasure to have spent the last ten years in academia, with all the joy and perks that come with it: living in Sweden and Australia, travelling the world, engaging with challenging ideas, and meeting countless inspirational people. Thank you for the ride. Susanna, it was your welcoming nature, inspiration, positive attitude, encouragement, and guidance that kept me going. Every PhD student needs a supervisor like you. Hugh, I highly appreciate and do not take for granted your close readings of my work and your critical but patient stance on the various avenues I have tried to pursue in this PhD project. I envy the clarity of your thought. My thanks also go out to Griffith University for providing me with my postgraduate research scholarships, and to the School of Humanities and the Safety Science Innovation Lab for hosting me.

I would also like to express my sincere gratitude to what I consider as my informal supervisory team: Eder, Betina, and JB. In our collaborations, over the years, we have had some academic fun 'kicking ass', discussing some great books, critiquing safety, and writing a number of interesting papers. Most importantly, however, you have become some of my best friends. Then there are Daniel and Jill, wise friends to share the ups and downs of PhD life with. Thank you all for your never ceasing support, I could and would not have done this without you. Thank you for bringing joy.

Of course, there is my eternal gratitude to my parents and brother, for their unconditional love, their encouragement to do whatever I want, and their support to go wherever I need to go—even if this means that your grandson is born on the other side of the world. Finally, I am especially indebted to my two favourite people in the world, my boy and his baby-mama. Thank you for giving me direction, without you I'd be lost. I love you above all.

Acknowledgement of published and unpublished papers included in this thesis

This thesis includes a mix of reformatted versions of sole-authored and co-authored published and unpublished papers.

Section 9.1 of the Griffith University Code for the Responsible Conduct of Research ("Criteria for Authorship"), in accordance with Section 5 of the Australian Code for the Responsible Conduct of Research, states:

To be named as an author, a researcher must have made a substantial scholarly contribution to the creative or scholarly work that constitutes the research output, and be able to take public responsibility for at least that part of the work they contributed. Attribution of authorship depends to some extent on the discipline and publisher policies, but in all cases, authorship must be based on substantial contributions in a combination of one or more of:

- conception and design of the research project
- analysis and interpretation of research data
- drafting or making significant parts of the creative or scholarly work or critically revising it so as to contribute significantly to the final output.

Section 9.3 of the Griffith University Code ("Responsibilities of Researchers"), in accordance with Section 5 of the Australian Code, states:

Researchers are expected to:

- Offer authorship to all people, including research trainees, who meet the criteria for authorship listed above, but only those people.
- accept or decline offers of authorship promptly in writing.
- Include in the list of authors only those who have accepted authorship
- Appoint one author to be the executive author to record authorship and manage correspondence about the work with the publisher and other interested parties.

- Acknowledge all those who have contributed to the research, facilities or materials but who do not qualify as authors, such as research assistants, technical staff, and advisors on cultural or community knowledge. Obtain written consent to name individuals.

Sole-authored paper

Included in this thesis is the paper in *Chapter 6* for which I am the sole author. The bibliographic details for this paper are:

Chapter 6 (submitted paper):

Winsen, R. van. (2014). *Situation awareness: The emergence of an object of knowledge*. Manuscript submitted for publication.

Co-authored papers

Additionally included in the thesis are papers in *chapters 2, 7, 8, 9* which are co-authored with other researchers. My contribution to each co-authored paper is outlined at the front of the relevant chapter. The bibliographic details for these papers, including all authors, are:

Chapter 2: (accepted paper)

Breakey, H., Winsen, R. van, & Dekker, S. W. A. (in press). 'Loss of situation awareness' by medical staff: Reflecting on the moral and legal status of a psychological concept. Journal of Law and Medicine.

Chapter 7: (published paper)

Winsen, R. van, Henriqson, E., Schuler, B., & Dekker, S. W. A. (2015). Situation Awareness: Some Conditions of Possibility. *Theoretical Issues in Ergonomics Science, 16,* 53-68.

Chapter 8: (published paper)

Henriqson, E., Schuler, B., Winsen, R. van, & Dekker, S. W. A. (2014). The constitution and effects of safety culture as an object in the discourse of accident prevention: a Foucauldian approach. *Safety Science, 11*, 465-476.

Chapter 9: (submitted paper)

Bergström, J., Winsen, R. van, & Henriqson, E. (2014). *On the rationale of resilience in the domain of safety: a literature review.* Manuscript submitted for publication.

The papers included in this thesis may differ slightly from their published versions. Adaptations had to be made in terms of page numbers, numbering of figures and tables, the layout of text, figures, and tables. Rather than at the end of each individual study, all references are listed in the reference list at the end of the thesis.

Other publications

Two other publications are not literally incorporated into the thesis like the above articles are. This is mostly due to considerations of the narrative and flow of this thesis. However, these publications, an opinion paper and book chapter, are directly informed by the research and writing conducted as part of this thesis. As their content overlaps with the *introduction, chapter 1, and the conclusion*, these publications are attached to the thesis as appendix A and B. The bibliographic details for these papers, including all authors, are:

Winsen, R. van, & Dekker, S. W. A. (in press). SA Anno 1995: A commitment to the 17th century. *Journal of Cognitive Engineering and Decision Making*.

Winsen, R. van, & Dekker, S. W. A. (in press). Human Factors and the ethics of explaining failure. In: S. Shorrock & C. Williams (Eds.). *Human Factors and Ergonomics in Practice*. Aldershot, UK: Ashgate Publishing Co.

Papers not included in this thesis

This thesis presents only a small portion of the larger and more diverse work of a three-year PhD project, in which I have collaborated in various research projects. During my PhD I have co-authored a number of other papers that have not been included in the narrative of the thesis, however, these papers are part of the larger PhD experience nonetheless. The bibliographic details for these papers, including all authors, are:

Henriqson, E., Bergstrom, J. Winsen, R. van, & Dekker, S. W. A. (2013). *Pilot-automation coordination breakdown in autothrust systems: a case study of automation surprise*. Manuscript submitted for publication.

Mayhew, B., Winsen, R. van, & Nyce, J. M. (2013). *New View Emergence in Accident Investigation*. Manuscript in preparation.

Schuler, B., Villela Pereira, M., Henriqson, E., Winsen, R. van, & Dekker, S. W. A. (2013). *Restorative Justice at School: The Effects of Power and Subjectivation*. Manuscript submitted for publication.

Appropriate acknowledgements of those who contributed to the research but did not qualify as authors are included in each paper.

(Signed) ______________________________ (Date)___________
Roel van Winsen

(Countersigned) ______________________ (Date)___________
Principal Supervisor: Sidney W. A. Dekker

(Countersigned) ______________________ (Date)___________
Principal Supervisor: Susanna Chamberlain

Introduction

Karl is in jail today. And the scientific disciplines of human factors and safety science might have had something to do with it. Karl Lilgert was the officer in charge of the large passenger ferry 'Queen of the North' when it ran aground and sank. Ninety-nine passengers and crew survived the sinking of the ferry. Unfortunately, two passengers were never found and presumed to have drowned. Despite Karl facing poor weather, substandard navigational equipment, inadequate company policies, and a lack of staff on the bridge, he was effectively blamed for the failure to turn the ferry in time (Keller, 2013). Karl was criminally prosecuted, particularly as the 'causes and contributing factors' in the accident investigation report strongly suggested that various distractions contributed to his failure to order the required course change (Transportation Safety Board of Canada, 2008). Additionally, the investigation report concluded:

> The working environment on the bridge of the Queen of the North was less than formal, and the accepted principles of navigation safety were not consistently or rigorously applied. Unsafe navigation practices persisted which, in this occurrence, contributed to the loss of situational awareness by the bridge team. (Transportation Safety Board of Canada, 2008)

As the court held that, "[m]aintaining situational awareness at all times and in all circumstances is key to proper navigation" (Supreme Court of British Columbia, 2013), the leap wasn't large[1]: the prosecution was able to successfully argue for the dereliction of his duty. Evidently, situational awareness was not maintained, otherwise the ferry would not have crashed and sunk. Karl was convicted for criminal negligence causing the deaths of the two passengers and sentenced to four years in prison. At the time I write these words, the appeal court has recently upheld his verdict, and Karl remains imprisoned (Keller, 2014).

[1] I deliberately simplify the complexity of the juridical case here, for the full verdict see Supreme Court of British Columbia (2013).

1. What?

What is the aim of this thesis? What are my research questions? What kind of analysis am I presenting? What are the objects of my analysis?

1.1 Aims & research questions

'Situation awareness', in Foucauldian terms, is an object of knowledge (Foucault, 2002a). As an object it is part of the larger discourses of human factors and safety science, together with other objects of knowledge like 'safety culture' and 'resilience engineering'. For Foucault, scientific objects are not the necessary or natural result of scientific progress, but rather the contingent effects of certain discursive practices of our scientific disciplines. In turn, objects have (discursive) effects: systems are designed to support operator situation awareness (SA), accidents are attributed to a loss of SA, corresponding training programs are set up, and procedures are introduced to address SA. Applied disciplines, such as human factors and safety science, however, typically do not provide much critical reflection on their discursive practices—on the interventions, solutions, classifications, and objects that they promote. This raises questions about whether such practices are having the effects that they aim to achieve.

The archaeological work of Michel Foucault explicitly sets out to understand how discourses, and the objects in them, emerge and discursively function. Foucault is not so much interested in the ontological or epistemological status of discourses and the object in them, as in what makes them possible. Surprisingly, the work of Foucault has been largely overlooked by the disciplines of human factors and safety science. This thesis presents a modest attempt to introduce (a part of) the work of Foucault into the domains of human factors and safety science. More specifically, this thesis provides a number of archaeologically inspired analyses of the discursive constitution and effects of some prominent objects of the human factors and safety sciences discourses: it chiefly focuses on situation awareness (SA) and later expands the scope by briefly looking at safety culture and resilience as objects of knowledge.

In doing this, this thesis asks the following question and series of sub-questions:

What insights can Foucault's archaeological analyses of objects provide the disciplines of human factors and safety science, and can these help in critically reflecting on some of the discursive practices that these disciplines promote?

- How does 'situation awareness' emerge and function in the scientific literature?
- What are the conditions of possibility for the emergence of 'situation awareness' as a ubiquitous object of knowledge?
- How is the object of 'safety culture' constituted and what are some of the effects of this discursive object?
- How are the object and associated subject of 'resilience' constituted and what are some of the effects of this discursive object?

1.2 Theoretical framework: the archaeological analysis of objects

This thesis first arose out of a passage in 'The Order of Things', out of the laughter that shattered, as I read the first page:

> This book first arose out of a passage in Borges, out of the laughter that shattered, as I read the passage, all the familiar landmarks of my thought – our thought, the thought that bears the stamp of our age and our geography – breaking up all the ordered surfaces and all the planes with which we are accustomed to tame the wild profusion of existing things, and continuing long afterwards to disturb and threaten with collapse our age-old distinction between the Same and the Other. This passage quotes a 'certain Chinese encyclopaedia' in which it is written that 'animals are divided into: (a) belonging to the Emperor, (b) embalmed, (c) tame, (d) sucking pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies'. In the wonderment of this taxonomy, the thing we apprehend in one great leap, the thing that, by means of the fable, is demonstrated as the exotic charm of another system of thought, is the limitation of our own, the stark impossibility of thinking that. (Foucault, 2002b, p. xvi)

The taxonomy Foucault provides from Borges, was, at one time, a meaningful and useful taxonomy. This is the point that Foucault wants to make in his archaeological projects: that our knowledges are historically contingent. The emergence of our body of scientific knowledge is not natural or necessary—not the logical end-point of scientific progress, which brings us ever closer to a 'true' understanding of the world—but just one possible way of understanding the world. Foucault's conception of knowledge is thus explicitly anti-teleological, as it does not involve any assumptions of progress.

In seeing 'discourse as practice,' Foucault (2002a) outlines the intentions for his archaeological approach, in which he regards discourses as systems of statements that emerge under historically and culturally specific 'rules' that determine what can meaningfully be thought and said at a particular time and place. Archaeology, then, is the study of the 'archive'—"the general system of the formation and transformation of statements" (p. 146). In his various archaeological projects, Foucault argues that the regularities we see in reality are brought about by the regularities of discourse, which we 'violently' impose on reality (Foucault, 1981). Discourse does not represent reality, but constitutes it: it brings it into view. Similarly, our objects of scientific knowledge are not 'out there' in the world waiting to be 'discovered'; they do not mirror some kind of external reality, but rather are the effects of certain historical discursive practices (Foucault 2002a).

This idea that our objects are historically contingent—and constructed—is something that does not receive a lot of attention in some scientific disciplines, particularly those with a more practical orientation such as human factors and safety science. Nikolas Rose (1996), a British (Foucauldian) sociologist, argues that disciplines such as psychology—and in this thesis I argue the same holds for human factors and safety science—suffer from a 'retrospective illusion' that prevents them from seeing that their objects are historically constituted. "[T]he object of psychology cannot be regarded as something given, independent, that pre-exists knowledge and which is merely 'discovered'. Psychology constitutes its object in the process of knowing it" (Rose, N. S., 1996, p. 109). Foucault, similarly, explains that the "psychiatric discourse finds a way of limiting its domain, of defining what it is talking about, of giving it the status of an object – and therefore of making it manifest, nameable, and describable" (Foucault, 2002a, p. 46). He maintains that the construction of categories and description of disorders serves to provide the human sciences with 'locatable objects of scrutiny' (Graham, 2011).

This constructionist argument does not lead Foucault to pose ontological or epistemological questions about our objects, as he is not interested in assessing the truth-value of our knowledges. Rather, Foucault wants to understand how objects are formed and attain the status of 'truth'—or as he calls it, how they enter 'in the true' (Foucault, 1981). Beyond the constitution of objects, Foucault explains that, "it's not just their formation that interests me, but the effects in the real to which they are linked." (1991b, p. 85). As such, he is interested in the effects of the objects that we construct (especially the ones that function under the guise of 'scientific truth'); that is, the consequences of constructing the world in a certain manner. In his genealogical projects Foucault will take this line of inquiry further as he analyses the effects of knowledge in terms of power and subjectivity. For now, however, Rose provides an eloquent quote that captures some the assumptions and intentions of Foucault's archaeological approach:

> Contemporary scientific reality and this goes for a science like psychology as much as any other is the outcome of the categories we use to think it, the techniques and procedures we use to evidence it, the statistical tools and modes of proof we use to justify it. But this does not amount to a delegitimation of its scientific pretensions. It is merely the basis from which we become able to pose questions concerning the means of construction of these new domains of objectivity and their consequences. (Rose, N. S., 1996, p. 109)

Archaeology thus provides us with a basis to start asking questions about the formation of (scientific) discourses and the effects that they have. This thesis follows this archaeological imperative: it aims to provide an archaeologically inspired analysis of a number of prominent human factors and safety science objects—situation awareness, safety culture, and resilience engineering.

1.3 Considerations of scope

Studying the discourses of human factors and safety sciences, even if only from a Foucauldian archaeological perspective (as opposed to Foucault's genealogical or ethical projects), would be outside the scope of a PhD thesis. Such an endeavour would probably constitute a life-long agenda for academic research. To make this project into a 'doable' PhD thesis, I have therefore limited the initial scope of this dissertation to the archaeological study of just one prominent object of the human factors and safety science discourses, that of 'situation awareness' (SA).

Originally adapted from its more commonsensical use in the air force—as an awareness of the situation, or specifically, the enemy—human factors and safety science picked up SA as a scientific object during the 1980's (Stanton, Chambers, & Piggott, 2001). However, it was not until the journal of Human Factors published its special issue on the topic in 1995 that the concept flourished as a scientific object. Currently, SA is one of the most prominent—that is, most widely published and cited—objects of the human factors discourse. When Lee, Cassano-Pinché, and Vicente (2005) examined the impact of papers published in the journal of Human Factors between 1970 and 2000, they found that four of the top ten most cited papers published (from 1990 to 1995) were on SA. A 2014 Scopus search confirms that this remains the case. SA thus constitutes an important object in the discourse of human factors and the associated discipline of safety science. "During the past 15 years, the concept of situation awareness has entered the mainstream of human factors, and the term has certainly entered the vernacular of human factors researchers" (Wickens, 2008, p. 397).

SA is not only a prominent object of the human factors discourse, but it can also be seen as representing the human factors discourse at large. According to a number of review papers on SA (Adams, Tenney, & Pew, 1995; Durso & Gronlund, 1999; Salmon et al., 2008; Stanton, 2006; Stanton et al., 2001; Stanton et al., 2006; Tenney & Pew, 2006), three models of SA dominate the literature: (1) Endsley's (1995b) three-level model of SA, which regards SA as the product of consecutive stages of 'information processing'; (2) the ecological model of SA, often based on Neisser's (1976) 'perceptual circle', and (3) the interactive subsystems model of SA, which is based on an 'activity theory' approach to cognition (as most prominently outlined by Bedny & Meister, 1999). These manners of specifying SA are representative of various theoretical mainstreams in the human factors discourse: "it was inevitable that some of the same arguments, such as those between the information-processing and ecological schools, would resurface regarding this new phenomenon" (Tenney & Pew, 2006, p. 3). Similarly, Durso and Gronlund regard SA as a good instance for the human factors discourse at large:

“Difficult to define, but easy to recognize, situation awareness is a test bed that can be used to determine if cognitive theory can ‘scale-up’” (1999, p. 305).

Initially, this thesis focuses on the analysis of the prominent human factors and safety science object of SA. As an archaeologically inspired study, the analysis is not interested in the value or epistemological status of SA—something that other studies have already extensively addressed (to name just a few, Billings, 1995; Dekker & Hollnagel, 2004; Dekker, Hummerdal, & Smith, 2010; Endsley, 2014; Flach, 1995b; Parasuraman, Sheridan, & Wickens, 2008; Salmon et al., 2007)—but rather in understanding the emergence of SA as a prominent object of knowledge. This thesis asks new questions about this object, and proposes a new kind of analysis: How does SA emerge in the literature and what discursive need does it aim to fulfil? What were the conditions of possibility of SA? The analysis that arises out of such archaeological questions will, hopefully, provide new insights into some of the discursive effects of SA, which, in turn, may lead us to ask new questions about our human factors and safety science practices at large.

The final part of this thesis will present two chapters that extend the initial scope of the project, both in terms of the object of analysis (pun intended) and the theoretical approach that they take. These final chapters take ‘safety culture’ and ‘resilience’ as discursive objects. Safety culture is an object that claims to function at the level of organisations, rather than that of individuals (as SA primarily does). In this respect, resilience engineering presents a particularly interesting object, as it is not always clear whether the subject of this object is the individual or the organisation. In order to start asking questions about the effects of these objects, these final chapters slightly expand the archaeological scope of the project, and start to draw on Foucault’s genealogical work to include questions about subjectivity and power. This last part also functions to show the wealth of possibilities that Foucault’s archaeological (and genealogical) approach provides.

2. Why?

Why apply Foucault’s archaeological approach to the disciplines of human factors and safety science? Why ask questions about the contingent nature of our discourses? Why describe the constitution and functioning of our objects of knowledge? Why might this be useful or relevant for human factors and safety science?

The reasons for my project are threefold. First, because Foucault's work, particularly his archaeological projects, has largely been overlooked by the disciplines of human factors and safety science, this thesis explores the application of (primarily) his archaeological approach to these domains of study. This is an innovative endeavour simply because it has not been done before and we thus do not know what new insights it may yield. Second, is the larger emancipatory project, which is entirely in line with Foucault's own reasons for his (historical) archaeological projects. By showing the contingency of our (scientific) knowledges and practices, we might be able to see that there is nothing necessary—or as Foucault would say 'natural'—about them. Showing the 'arbitrariness' of our human factors and safety science knowledges will hopefully open up possibilities for the examination of some of their taken-for-granted truths. Third, but possibly most important, is the more specific reason to provide some reflection on certain human factors and safety science practices. As applied sciences, these disciplines show little critical reflection on some of their practices. This lack of reflection, consequently, may mean that some of those practices have effects that depart from the disciplines' explicit aims.

2.1 Introducing Foucault to the disciplines of human factors and safety science

The work of Michel Foucault, arguably one of the most influential social scientists of the twentieth century, has largely been overlooked by the disciplines of human factors and safety science. Foucault's ideas and criticisms have had a large impact on a wide range of scientific disciplines, from sociology and psychology to linguistics and history. It is therefore remarkable that social science disciplines as human factors and safety science have neglected Foucault, especially as these disciplines blossomed in the late seventies and eighties—in the wake of dramatic accidents as the Tenerife aviation disaster (1977) and the partial nuclear meltdown at Three Mile Island (1979). This period is also the time in which Michel Foucault wrote his celebrated books and was a prominent social figure (both inside and outside France). Since Foucault's death in 1984, the relevance and popularity of his ideas have little declined. The fields of human factors and safety science have in the meantime continued to steadily grow under the impetus of the rapidly increasing complexity of our modern social-technological systems. However, thirty years after the death of Foucault, human factors and safety science still seem to avoid Foucault's critiques and suggestions.

Besides the complexity and density that make Foucault's form of analysis less accessible, particularly to those in more practice-oriented fields (Graham, 2011), one reason for why the applied disciplines of human factors and safety science have overlooked Foucault's work may lie in the difficulty to see its practical yield. This could also explain why Foucault's genealogical approach—which relates its conception of discourse and knowledges to that of power, subjectivity and society— has been more popular and widely used than his archaeological approach—which presents a narrower, more theoretical and formative conception of discourse. Apart from a few unique papers (Almklov, Rosness, & Størkersen, 2014; Antonsen, 2009; Rasmussen, 2010, 2013; Turner, 1989), most of the studies that fall within the scope of human factors or safety science and that do explicitly draw on a Foucauldian framework for analysis come from the domain of nursing (e.g. Holmes, Murray, Perron, & McCabe, 2008; Holmes, Murray, Perron, & Rail, 2006; Murray, Holmes, Perron, & Rail, 2007; Murray, Holmes, & Rail, 2008). These papers largely draw on Foucault's genealogical approach, especially in terms of 'regimes of truth', as it provides them with a manner to examine the power relations in their associated fields of work.

Organisational studies—which presents a discipline (or cross-disciplinary research orientation) that not only shares some of human factors and safety science' concerns but also resembles these disciplines in regards to its practical orientation—seems to have embraced the work of Foucault on a larger scale. Here, again, the genealogical approach is far more popular than the archaeological (e.g. Barker & Cheney, 1994; Bergström, Hasselbladh, Kärreman, & La Porte, 2009; Bergström & Knights, 2006; Chan, 2000; Clegg, 1989; Esbester, 2008; Kärreman et al., 2009; Knights & McCabe, 2003; McKinlay & Starkey, 1998; Seeck & Kantola, 2009), but extensive literature searches did unearth a few papers that explicitly present an archaeological approach to studying discourse (Topp, 2000; Xu, 2000). Unfortunately, these papers do not engage significantly with topics of interest to human factors and safety science. Rasmussen's work (e.g. Rasmussen, 2010, 2013) presents one of the few exceptions, but, as stated, his work relies on a genealogical, rather than archaeological approach.

The lack of reference to Foucault, especially to his archaeological work, that human factors and safety science display, cannot be explained by saying that the topics of Foucault's analysis are not of interest to these disciplines. Foucault himself, throughout his career, encouraged his readers not to slavishly follow his analysis, on whatever topic, at a literal level, but rather to draw on his work as a resource or 'tool' for thinking and questioning in general. With his work, Foucault hopes to inspire his readers to question the solidity of their own knowledges. In an interview published in Le Monde he states:

> … a book is made to be used in ways not defined by its writer. The more, new, possible or unexpected uses there are, the happier I shall be … All my books are little tool-boxes. If people want to open them, to use this sentence or that idea as a screwdriver or spanner to short-circuit, discredit systems of power, including eventually those from which my books have emerged … (Foucault, cited in Patton, 1979, p. 115)

This thesis is inspired by Foucault's call to question our contemporary knowledges—which is a prime archaeological goal. It presents a modest, but rather unique, attempt to introduce Foucault's archaeological approach to the scientific disciplines of human factors and safety science. It does so by using and developing Foucault's archaeological work as a 'tool' to describe the discursive constitution and effects of some objects that feature prominently in the human factors and safety science discourses.

By introducing a Foucauldian—especially archaeological—analytic to the fields of human factors and safety science, this thesis hopes to provide them with new ways of reflecting on our knowledges and practices. It hopes to bring diversity to the social scientific approaches that these disciplines typically rely on. It does so "by opening up alternative ways of investigating knowledges, subjects traditional research methods to scrutiny, and raises new questions concerning … research. All this acts to enliven and enrich social … research and practice by opening up, rather closing off, debate" (Garrity, 2010, p. 207).

2.2 Foucault's emancipatory project

In 'The Archaeology of Knowledge,' Foucault (2002a, p. 205) explains: "To tackle the ideological function of a science (or of a discourse with scientific pretentions) in order to reveal and modify it," you should "question it as a discursive formation; it is to tackle not the formal contradictions of its propositions, but the system of formation of its objects, its types of enunciation, its concepts, its theoretical choices". Ultimately, Foucault's archaeological project is emancipatory. Through understanding how particular conceptions of the world emerge, we can see the contingency and arbitrariness of our knowledges, which opens up possibilities for the examination of certain taken-for-granted truths. Mills explains this forcefully:

> This is why Foucault's archaeological analysis of discourse is important; he is not interested in simply analysing the discourses which are circulating in our society at present, what he wants us to see is the arbitrariness of this range of discourses, the strangeness of those discourses, in spite of their familiarity. He also wants to chart the development of certain discursive practices, so that we can see that, rather than being permanent, as their familiarity would suggest, discourses are constantly changing and their origins can be traced to certain key shifts in history (Mills, 1997, p. 26).

In short, with his archaeological projects Foucault aims to render the familiar strange, to display the arbitrariness of some of our knowledges, discourses, and truth. Whereas the realisation that our knowledges, in particular our scientific knowledges and objects, are arbitrary, may be a matter of despair for some people, "Foucault sees it as a reason to be optimistic. Ideas and practices which have oppressive and unjust effects on people and limiting effects within knowledge and science can always be changed" (O'Farrell, 2006, p. 189).

By questioning our taken-for-granted (scientific) truths, by showing their arbitrariness, we can change our (discursive) practices and knowledge. This is the emancipatory agenda that this thesis aims to pursue: by asking questions about the historically contingent nature of some of our prominent human factors and safety science objects, it aims to open up possibilities for new or oppressed knowledges and truths—for new articulations and specifications of objects.

Not only may Foucault's archaeological approach help to challenge some (dominant) discourses that are taken as self-evident, but by introducing Foucault's archaeological approach to the disciplines of human factors and safety science, I also hope to add to the theoretical and methodological pluralism—or as Healy (2003) so elegantly labels it, the ‘epistemological pluralism’—of these domains. The ‘repertoire’ of methods and epistemologies that these applied disciplines typically draw upon is dominated by those of engineering and experimental psychology (see chapter 1). A few years ago, Batteau (2001, 2002) convincingly argued for an anthropological approach to expand the ‘tool-box’ of human factors and safety science approaches:

> The anthropological approach described here should be seen neither as an alternative nor a negation of the approaches that have dominated Human Factors studies for the past sixty years. Instead, anthropology can add another dimension to the core approaches that have dominated Human Factors. (Batteau, 2002, p. 166)

In line with Batteau’s plea for the inclusion of anthropological inquiry, this thesis asks if Foucault’s archaeological approach can further add to the ‘repertoire’ of approaches that human factors and safety science draw upon. As Gutting explains about Foucault’s archaeologies: “by unearthing alternative ways of thinking and acting, they end the de facto monopoly of the dominant system” (1989, p. 107).

2.3 Reflecting on the disciplines of human factors and safety science

Human factors and safety science are disciplines that take pride in their practical orientation. Early pioneers of these disciplines, such as Fitts and Jones, were experimental psychologists who adapted their laboratory techniques to applied problems, and in doing so created an important new way of analysing accidents and human performance issues. They introduced psychology’s empirical-experimental methodology to the study of ‘real world’ complex social-technical problems (see chapter 1). Even though some human factors and safety science researchers have recently objected to the distinction of applied versus basic science (Hancock & Drury, 2011; Helton & Kemp, 2011), generally these disciplines are happy to be classified as applied: “Applied science would be science with the expectation of a relatively short duration of work before a practical return on the invested effort is obtained, e.g. research geared to a relatively specific and immediate issue at hand” (Helton & Kemp, 2011, p. 400). An applied orientation, however—by its definition—does not leave much time for the critical reflection on the practices, recommendations, categories, or objects for analysis that applied disciplines promote.

Of course, in their 'normal scientific' pursuits (Kuhn, 1996), human factors and safety science do assess the results (effects) of their interventions and recommendations. This, after all, is at the core of the experimental method (or more specifically, the hypothetico-deductive method), which is characteristic of human factors and (perhaps to a slightly smaller extent) safety science.

The scientific literatures of human factors and safety science provide a number of publications that present reflections on the epistemology or ontology of our constructs and explanations (Aven, 2014; Dekker, Nyce, Winsen, & Henriqson, 2010; Dekker & Hollnagel, 2004; Haavik, 2014; Wilson, 2000). Similarly, there are several publications that ask questions about the ethics and (ultimate) goals of these disciplines and some of their practices (e.g. Cook & Nemeth, 2010; Corlett, 2000; Dekker, 2014; Hancock & Drury, 2011). However, in searching for publications that explicitly study the discursive formation of the 'products'—solutions, categorisations, interventions, objects—of human factors and safety science, I struggle to find publications that reflect on the discursive practices promoted by these disciplines.

It is vitally important for any science to critically reflect on the knowledge and interventions it proposes. Hollnagel puts it slightly milder by stating, "it behoves every science every now and then to look at itself in a critical manner, to assess the achievements of the past, and to consider where it should be going in the future" (Hollnagel, 2014a, p. 40). Similarly, the editors of the flagship journal 'Safety Science' explain about their call for critical and reflective papers on the discipline of safety science: "We believe that it is important for the wellbeing of a scientific discipline that theories and methods are examined every so often and internal investigations may also come to represent starting blocks for the creation of new ideas" (Le Coze, Pettersen, & Reiman, 2014, p. 1). It is thus important for the success of human factors and safety science as scientific disciplines, that they are able to take a critical stance on, and be able to question, the categories (objects), theories, and interventions that their fields promote.

Not only is such reflection theoretically important, it is also of practical and ethical importance to look at—and take responsibility for—the effects of our scientific discursive practices. "In the case of sciences whose *raison d'être* is their practical use rather than the neatness of their theories – and HFE is an example of that – then the achievements and the future direction are of particular importance" (Hollnagel, 2014a, p. 40). Corlett (2000, p. 679), in his article on ergonomics and ethics, explains this beautifully: "It is because what we do has an influence on what happens to our world. Our business is with people, helping to make their world one in which they can operate effectively". Now go back to the case of Karl Lilgert, who was prosecuted for criminal negligence, with which I started this introduction. The case caught my attention a few years ago and now runs through this thesis as the 'red thread' that draws the whole together. It does so, because (for me) it so clearly shows the need for reflection on our practices.

Karl's case presents an interesting paradox. According to the International Ergonomics Association (2014), human factors is "the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance." With these goals—system efficiency and human wellbeing—in mind, the scientific disciplines of human factors and safety science embraced SA as a discursive object to aid our understanding of human decision making in complex dynamic systems and help with the design of human-machine interfaces. In the case of Karl, however, SA—and the loss of it—functions as a causal and normative construct to explain and judge (allegedly poor) human performance. As such, we need to ask ourselves whether the object of SA may have effects that run counter to the disciplines' explicit aim of optimising system efficiency and human well-being.

Critical reflection on the effects of our discursive practices is thus both necessary in terms of the livelihood and theoretical innovation of any scientific discipline, as well as for the practitioners and larger society that suffers the practical effects of their propositions. Foucault, in a chapter that is appropriately entitled 'So is it important to think?' specifies what such critical reflection could look like: "A critique does not consist in saying that things aren't good the way they are. It consists in seeing on what type of assumptions, of familiar notions, of established, unexamined ways of thinking accepted practices are based" (Foucault, 2001, p. 456).

This thesis will explore whether Foucault's archaeological approach can provide this kind of reflection on the practices that the disciplines of human factors and safety science promote. In archaeologically analysing the rules, relations, and 'conditions of possibility' that govern the formation of objects within the discourses of human factors and safety science, it aims to see 'on what type of assumptions, of familiar notions, of established, unexamined ways of thinking accepted practices are based'. As such, this thesis aims to, first, see what insights an archaeological analysis of some prominent human factors and safety science objects can provide, and, second, to draw on these archaeological insights to reflect on some of the discursive practices that these disciplines are promoting.

3. How?

How will I pursue these aims? How is this thesis presented? How am I going to organise my thesis? How is the reader supposed to read this manuscript?

3.1 Methodological approach

How am I going to approach the goals of this thesis methodologically? This might be the largest problem to solve in this thesis. As this thesis asks what insights an archaeological approach can provide disciplines of human factors and safety science in order to reflect on some of its practices, it is one of the aims of this thesis to explore how we can use—or even 'translate'—Foucault's archaeological approach in these disciplines.

Because this thesis presents one of the first attempts to bring Foucault, especially his archaeological approach, to the disciplines of human factors and safety science, there is no available 'template' to follow. Nor does Foucault himself provide a prescriptive, mechanical, method to 'do' archaeological analysis. Through studying and reviewing Foucault's elaborate body of archaeological work, a large part of this thesis is directed at teasing out a number of analytical directions—or better, analytical concepts—that will be suitable for this thesis' purposes. As such, this thesis will not present a 'full' archaeology. Rather, it will present a number of analytical studies that are inspired by Foucault's archaeological approach and aim to study a number of prominent human factors and safety science objects at an archaeological level. The analytic and methodological approach will find its final form in the various archaeologically inspired studies presented in this thesis.

3.2 Thesis as a collection of published and unpublished papers

This thesis is organised as a series of papers. “This thesis format may include one or more papers that have been prepared, submitted, or accepted for publication. A Doctoral or MPhil thesis prepared in this way is not a different degree.” (Griffith University, 2014). The Griffith University website presents a list of advantages to organising a thesis in this manner. Not only will it save time and effort to prepare chapters for publication when they are already written as journal articles, it is also a good way to disseminate your work—in a more timely manner than waiting for people to pick up your monograph—receive peer-reviewed feedback on your work (which may increase the quality), and may provide your examiners with the confidence that you know what you are doing, academically (Griffith University, 2014).

However, there is also a cost to presenting a thesis, as a whole, in this manner.
The downside of writing journal articles, instead of ‘regular’ monograph chapters, is that they have their own local constraints and requirements: as each article needs to present a stand-alone argument, each article needs its own introduction, theoretical framework, methodological guidelines, etc. As such, when reading this thesis as a whole, there will necessarily be some overlap or redundancy. Moreover, the goal of a specific article may not precisely align with the role that a chapter in a monograph needs to fulfil. Similarly, as the audience of the article may differ from that of the thesis, there can arise conflicts in the chosen vocabulary, references, writing style, and even length of the argument.

The article that constitutes most of chapter 2, for example, is written with a different intent (and audience) than the role that that it now needs to fulfil in the thesis. The article was born out of my frustration with Karl’s prosecution in particular, and the role that human factors and safety science objects, such as SA, play in the process of criminalizing error in general (this latter motivation overlaps largely with my aims in this thesis). As such, the article argues for the inappropriateness of using SA—or ‘loss of SA’—in this specific manner. For short-term publication reasons, the co-authors and I chose to pitch our argument to a medical audience and aim for publication in a medical journal (with which come strict restrictions for the length of our manuscript). Chapter 2, as part of the thesis, however, primarily functions to show that some of the discursive objects that human factors and safety science propose, may function counter to the explicit goals of these disciplines. In light of the larger thesis, chapter 2 aims to show the need or relevance for reflecting on our discursive practices. This is the 'price' that needs to be paid for pursuing a PhD as a series of papers as opposed to writing a more coherent monograph.

Overall, as it is this thesis' goal to bring Foucault to the disciplines of human factors and safety science, the benefits of pursuing a 'thesis as a series of published and unpublished papers' (Griffith University, 2014) outweigh the costs outlined above. Peer-reviewed journal articles, as opposed to chapters in a monograph, have a (much) higher chance of being read by human factors and safety science scholars. Also, because the peer-review process functions as an apparatus of quality control, the arguments obtain a higher 'status' as academic documents. As such, during my thesis I have aimed to write, as much as practically possible, journal articles instead of 'regular' chapters. Besides the analytical studies that are all written as journal articles, I have been able to rewrite parts of the introduction, chapter 1, and the conclusion into manuscripts that have been accepted as an opinion paper and a book chapter in an edited book (see appendix A and B).

All the papers included in the thesis have been submitted for publication in either human factors or safety science journals. Some papers have been published by the time that I am writing this, others have just been accepted, and two papers have just been submitted and await peer-review. As multi-authored papers are normal in the disciplines of human factors and safety science, I am a co-author on a number of papers. To allow these papers or manuscripts to be included in this thesis, I am required to have made a 'significant' contribution to the paper (Griffith University, 2014). Whereas in the majority of the papers I have had a leading role in writing the paper and facilitating the discussions with the other authors, in others I have been a member of the author team and I have taken my turn in the various iterations that typically make up the various drafts of the manuscripts. In these turns, I have contributed by writing parts of the arguments, critically revising the work of the other authors, editing the text and lay-out, etc. My contribution to a specific paper is stated at the beginning of each chapter that includes a co-authored paper.

3.3 Organisation of thesis

For reasons of transparency, this thesis is organised into four parts. Each part, consisting of a number of chapters, has its own role to play in the thesis, which will be elaborated upon hereunder.

Part I – Human factors and safety science & the problematic use of situation awareness

Part I presents a relatively short section that aims to set the stage and relevance for reflecting on our human factors and safety science practices. Chapter 1 intends to concisely sketch the objectives of human factors and safety science as academic disciplines. By describing their (shared) goals, some of their central assumptions, and their applied nature, it aims to provide the foundation for later reflection on whether their objects are discursively functioning in line with their explicit aims. Chapter 2 zooms in drastically, and focuses on situation awareness (SA) as a discursive object that prominently appears in both discourses and constitutes the main object of analysis for this thesis. By outlining one particularly problematic manner in which the object is currently used—in legal discourses assessing culpability and criminal liability—this chapter lays out the (practical) relevance of the entire thesis. As a strong example, inspired by the case of Karl, chapter 2 places on full display that without reflection on the discursive use of our objects, they may start to function in ways that undermine the intentions with which they were conceived.

Part II – The theoretical and methodological framework for analysis

Part II presents a much larger and theoretically heavier section. Its main goal is to outline the theoretical and methodological framework for the analytical studies that will be presented in the later parts of the thesis. Chapters 3 and 4 provide elaborate descriptions of Foucault's intentions with, respectively, his notions of discourse and that of archaeology. As Foucault does not provide an instrumental method for archaeological analysis, the goal of these chapters is to draw out and clarify some of the assumptions underlying this kind of Foucauldian analysis, as well as to 'distil' a number of possible ventures for the archaeological study of objects. As such, these chapters introduce a number of 'analytical concepts' that can be used to inform my specific analyses. In making an effort to clearly delineate where these concepts for analysis come from, chapters 3 and 4 are rather general in their descriptions, so it is not until chapter 5 that I provide a number of methodological reflections on what this means for the analytical studies in this thesis. Chapter 5 provides an elaborate description of the methodological choices I have made in order to develop Foucault's archaeological theoretical framework for my specific purposes.

PART III – The archaeological analysis of situation awareness

Part III is where the archaeologically inspired studies into the object of SA find their final form. Together with Part IV this part constitutes the analytical core of this thesis. It is here that the overarching thesis question as to what insights an archaeological approach can bring to the disciplines of human factors and safety science finds some of its answers. The two chapters in this part of the thesis present two different angles—drawing on two different analytical concepts—of Foucault's archaeological approach. By drawing on Foucault's suggestions for 'rules of formation' of objects, chapter 6 provides a systematic analysis of the top-cited literature on SA to understand how this object emerges in the literature. Chapter 7, however, asks about the 'conditions of possibility' for SA, and presents my narrative about the (historically contingent) conditions that allowed SA to become a ubiquitous object of knowledge. Both chapters conclude by reflecting on the contingent nature of this discursive object and raise a number of questions about its effects.

PART IV – Extending the scope of analysis

Part IV functions to extend the scope of my analysis, both in terms of the objects of analysis and the theoretical and methodological approach that these final chapters take. Chapters 8 and 9 each present an archaeological analysis of a prominent safety science object, respectively that of 'safety culture' and 'resilience'; however they do so in quite distinctive manners. Chapter 8 provides a dense archaeological description of the constitution of safety culture as an object, after which it presents some effects of the object from a genealogical perspective. Alternatively, chapter 9 presents a study that is interested in the rationale and discursive constitution of the object and subject of resilience. It does so to ask a number of (ethical) questions about the discursive functioning of this object. The reason for including these two articles as chapters in this thesis is to display the broad applicability of the Foucauldian approach that I advocate.

Foucault's archaeological analysis does not only apply to objects that have their roots in the social sciences—what Foucault calls 'the sciences of man' (Foucault, 2002b)—but is equally appropriate to analyse objects that claim to function at the level of organisations. Moreover, Part IV hopes to show that archaeological analysis flows seamlessly into a Foucauldian genealogical analysis that focuses more on the effects of our knowledges in terms of the power relations and subjectivities.

Finally, in the conclusion I will reflect on the insights that these archaeological investigations provide the disciplines of human factors and safety science, as well as ask if these can help to critically reflect on some of their discursive practices. I will also briefly discuss the limitations of archaeological analysis, by reflecting on the kind of questions that it does not answer. Possibly, these are the questions where a genealogical perspective may take up where archaeology left off.

3.4 A short guide to the reader

As this thesis aims to bring Foucault to the disciplines of human factors and safety science, my principal audience is that of human factors and safety science. As such, readers already familiar with Foucault's archaeological work should feel free to skip chapters 3 and 4, as this part of the thesis is mostly about 'getting a grip' on some of Foucault's central analytical concepts for archaeologically studying objects. As a gesture to these readers, chapter 5 presents some of my methodological choices against a backdrop of Foucault's assumptions, and can be read without disrupting the flow of the narrative too much.

Moreover, as the analytical chapters are all written as journal articles, they each provide brief explanations of the theoretical (and methodological) framework used for their specific archaeological analyses. Consequently, readers only interested in seeing what specific insights an archaeological approach can provide—as this is the main aim of this study—can skip directly to Parts III and IV.

Similarly, readers already familiar with the disciplines of human factors and safety science can skip chapter 1, which aims to briefly sketch the goals for these disciplines. Chapter 2, however, may provide an innovative manner of critically reflecting on the (problematic) discursive use of a prominent human factors and safety science object—that of situation awareness.

PART I

\-

HUMAN FACTORS AND SAFETY SCIENCE

Chapter 1.
The objectives of human factors and safety science

Abstract

By sketching the historical imperative of human factors and safety science, this chapter aims to outline some of the goals and perspectives that characterise these disciplines. As scientific disciplines concerned with the safe and efficient operation of our socio-technological systems and the wellbeing of the people that work in them, human factors and safety science have rapidly developed in the 20th century. The disciplines share a historical trajectory in which they have moved from a reductionist focus that searches for component failures—most prominently human error—towards a systems approach in which accidents are seen as problems that can emerge even when all parts are functioning normally. Rather than seeing human error as inherent to human operators, from a systems approach, human performance is to be understood as part of the larger operational and organisational environments people work in. As such, both human factors and safety science draw on a systems approach to characterise and strengthen their status as scientific disciplines.

1.1 Scientific disciplines

To paraphrase Foucault (see 2002a, p. 152), it is not easy to characterise disciplines like human factors and safety science. They are characterised by unclear borders, ambiguous objects of study, a pragmatic lack in consistency in regards to a preferred method, and epistemologies lacking in reflection. Safety as a scientific theme interacts with many disciplines—engineering; psychology; sociology; anthropology; medicine; aeronautical, maritime, or other transportation studies; and environmental and organisational studies to name just a few. As such, safety science is often regarded a multi- or inter-disciplinary field. It is often assumed that progress in any of these disciplines coincides with progress in safety (Guarnieri, 1992). After all, newer better designs in engineering have to be safer as well. Leveson, however, argues: “Safety, like any quality, must be built into the system design" (2004, p. 252). As such, safety is more than the by-product of progress in other disciplines, and safety science can be considered an academic field or discipline in its own right (Hollnagel, 2014c). In his appropriately entitled paper, 'What is safety science?' Aven explains that, "there is only one field that specifically addresses the issue of developing suitable tools for understanding, assessing, characterising, communicating and managing safety, and that is safety science…" (2014, p. 18).

Similar discussions regarding the discipline of human factors and ergonomics (HFE), abound. Meister (1999), for example, explains that many human factors professionals feel what they do is part of psychology, engineering, or physiology. As such, he asks whether HFE is a subspecies of other disciplines; an interdisciplinary discipline; or a distinctive, hence autonomous, discipline that is different from other disciplines? Meister eventually concludes that no other discipline is concerned with the relationship between humans and technology: "the three major disciplines that have influenced HFE do not account for certain HFE aspects, then logically HFE must be a distinctive discipline" (Meister, 1999, p. 26). Wilson agrees: "Ergonomics does have a unique, defined place amongst the modern disciplines of study and practice: it is the discipline that provides the inter- and multidisciplinary theoretical understanding of all interactions in human-technology systems and application of such understanding in design" (Wilson, 2000, p. 566).

Even though some people like to distinguish between ergonomics, which emphasizes the physical aspects of work—anthropometry, workplace and tool design, etc.—and human factors, which emphasizes the cognitive or social aspects of work—stress, fatigue, team work, etc.—overall, human factors and ergonomics (HFE) studies the interaction of humans and (technological) systems (Meister, 1999). In my thesis, I will use 'human factors' to refer to this overall understanding of HFE (which thus includes ergonomics).

It would make for a good Foucauldian project to assess the unity, and thus the individuality of human factors and safety science as 'discursive formations'—Foucault asks: "What are these unities? … Are they merely a retrospective regrouping by which the contemporary science deceive themselves as to their own past" (Foucault, 2002a, pp. 34-35). However, it is beyond the aim and scope of this thesis to do so. In referring to human factors and safety science, I will simply adopt the perspective that regards them as distinctive disciplines, as both have features (orientation, methods, type of interventions, epistemological preferences, etc.) that are unique and differentiate them from other disciplines. As such, they can also be differentiated from one another. However, in this thesis I am not interested in exploring the idiosyncrasies of these two disciplines, rather I am interested in some of the features that they share.

In line with Wilson's (2000) characterisation of human factors above, according to the International Ergonomics Association (2014), human factors (and ergonomics) is "the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance." As accidents may jeopardise both the productivity and the wellbeing of those involved in or dependent on our socio-technological systems, safety is of central concern to human factors. This, obviously, is one of the main features it shares with safety science, and, as such, one of the central themes of this thesis.

Safety, however, is an 'uncertain object' (Foucault, 2002a). Many different definitions of safety exist. Often it is defined as the absence of negatives. Leveson, for example, defines safety as "the freedom from accidents (loss events)" (2012, p. 467). Safety is also typically associated with low or acceptable risk: the lower the risk, the higher safety (Aven, 2014). Safety is thus defined as the absence of negatives, as a 'non-event' (Hale, 2014; Weick, 2001): "the subject matter of safety science is therefore the occurrence—or rather, the non-occurrence—of adverse outcomes (accidents, incidents, and near misses) and their aetiology, but not safety as such" (Hollnagel, 2014c, p. 22). The question, then, becomes whether the absence of something offers a suitable object of study (see Hollnagel, 2014c).

Hollnagel provides a way out: "The solution is instead to change the definition of safety so that the focus is on what goes right rather than on what goes wrong" (2014c, p. 23). Instead of defining safety as the absence of negatives, Hollnagel points to ensuring things go right, or in other words, to ensure efficiency and human wellbeing (Hollnagel, 2014b). "The object of safety science is accordingly how people are able to provide the required performance under expected and unexpected conditions alike" (Hollnagel, 2014c, p. 24). The editors of the discipline's flagship journal, 'Safety Science', explain: "Safety as an object of scientific investigation can be claimed to have emerged in relation to social ambitions for increased protection" (Le Coze et al., 2014, p. 1). With these (positive) definitions of safety, the explicit orientations of safety science and human factors further align: both disciplines aim to optimise system efficiency (which includes avoiding accidents and striving for success) with the wellbeing of the people in and dependent on our socio-technological systems in mind.

In sketching the (historical) imperative of these disciplines, this chapter aims to further outline some of the characteristics that the disciplines of human factors and safety science share. As the larger thesis aims to explore if Foucault's archaeological approach can provide insights and reflections on some of the discursive practices that these disciplines promote, the goal of this chapter is to get a grip on the objectives and perspectives of these disciplines. Later, partly in the individual studies and partly in the conclusion of this thesis, this will provide the basis for reflection on the question whether some of the discursive practices that human factors and safety science promote—situation awareness, safety culture, and resilience—are functioning in a manner that runs counter to the goals that these disciplines have.

1.2 The birth of a safety science

Throughout human history, accidents have occurred, and for the majority of this history, accidents have been regarded as truly 'accidental', meaning random and unexplainable coincidences, sometimes referred to as 'acts of God'[2] (Loimer & Guarnieri, 1996). However, in the beginning of the nineteenth century, with the rise of the industrial society and declining mortality rates from infectious diseases, accidents became a health issue as well as an economic problem. With the industrial revolution at its height, an increasing number of miners, factory workers and other equipment operators were exposed to dangerous new machines and working conditions. With these vast societal changes came an increase in accidents that could be attributed to the new technology (Coury, Ellingstad, & Kolly, 2010; Guarnieri, 1992).

[2] There is an interesting paradox here, as there is a difference in seeing accidents as 'random events' or seeing them as 'acts of God'. The latter reference displays the necessity to provide some kind of 'explanation' for why bad things happen, which thus opposes the idea of seeing accidents as random 'accidental' events.

Although (reliable) statistics were not available until the start of the twentieth century, by the 1850s alcohol consumption by operators was regarded as one of the main causes of accidents (Guarnieri, 1992). This idea of alcohol causing accidents represented a move away from the 'accidental'—random acts (of God)—meaning of accidents, towards the enlightenment idea that people are responsible for their own actions and have to bear the consequences of them. For safety, this meant that people were now considered responsible for their own safety (Guarnieri, 1992). As such, early approaches to accident investigation, carried out as a coroner's inquest, approached accidents as criminal investigations and set out to allocate blame and determine who was responsible for the harm (Coury et al., 2010).

At the end of the nineteenth century, this 'mechanism of responsibility' was replaced by a different arrangement, that of solidarity. Rather than being based on fault and risk, the main instrument of the 'mechanism of solidarity' is insurance (Ewald, 2002). Sponsored by workers insurance companies, and due to the rapidly increasing number of work related accidents in all industrial societies, the start of the twentieth century saw the first systematic research efforts into the causes of accidents. This research revolved around two believes: first, that safety is cost effective, and second, that accidents had psychological causes. Psychologist argued that accidents were caused by mental errors as attitude, nervousness, worry, and fear. These ideas were further supported and reinforced by the rise of Freudian theories of human behaviour (Guarnieri, 1992). In practice, supervisors were taught to diagnose problems and treat employees by talking about the occurrences they were involved in (Heinrich, 1941).

This psychological approach generated many research opportunities, and was consequently happily accepted by the academic community. However, during the 1930s this enthusiasm had largely disappeared, as it had failed to deliver evidence for its increase in safety—it had failed to transfer its ideas to practical safety programs and measurable safety benefits (Guarnieri, 1992).

Alternatively, in his seminal book 'Industrial Accident Prevention' from 1931, Heinrich (1941) introduced the idea that accidents might be the result of a 'sequence of events', rather than being caused by single (human) failures. This sequence of events, with the accident—or as Heinrich terms it, the 'preventable injury'—being the final event, results from human interaction with a 'hazard'.

> The occurrence of a preventable injury is the natural culmination of a series of events or circumstances, which invariably occur in a fixed and logical order. One is dependent on another and one follows because of another, thus constituting a chain that may be compared with a row of dominoes placed on end and in such alignment in relation to one another that the fall of the first domino precipitates the fall of the entire row. An accident is merely one link in the chain. (Heinrich, 1941, p. 14)

Heinrich (1941) thus started to move away from the psychological approach, which puts the responsibility of the entire accident on the shoulders of the human, and suggested an accident model that represents accidents as a sequence of factors that are linked together like a chain of events, with interdependencies among these factors.

An important principle in Heinrich's 'domino model' of accidents is that the removal of a critical factor in the sequence of events would prevent the accident from occurring. "In accident prevention the bull's-eye of the target is in the middle of the sequence—an unsafe act of a person or a mechanical or physical hazard" (Heinrich, Petersen, & Roos, 1980, p. 22). Consequently, according to Heinrich (1941), 88 per cent of accidents are caused by unsafe acts of persons and 10 per cent by unsafe machines (which leaves just 2 per cent of accidents as being unavoidable). This left Heinrich to formulate his second 'axiom', which states that: "The unsafe acts of persons are responsible for a majority of accidents" (Heinrich, 1941, p. 12). As such, Heinrich's main focus—or locus for intervention—is still on the individual human operator.

1.3 Towards a human factors approach

During the twentieth century, we have seen a slow move away from attributing errors and accidents solely to humans, towards looking at how systems, in which these errors take place, shape human behaviour. Prior to World War II, trial and error was the only selection mechanism to match humans with technology, machines, and tools; the operator was either able to work with the machine, or he was not. This 'selection' process continued until a suitable person was found. The point here is that the human was to fit the tools or adjust to the machine, and that no efforts were made to design the machine or tool to fit the needs of the human. As an example, in the military, which has always been at the forefront of human factors research, anyone small enough to fit in a tank automatically became a tank operator (Meister, 1999).

In the beginning of the twentieth century, selection mechanisms started to slowly develop and include other factors than people's physical size. Criteria for particular capabilities came into play, such as a person's ability to cope with a lack of oxygen, anxiety of being locked up in small spaces for long periods of time, and other stress related factors (Meister, 1999). World War I stimulated the rapid development of more sophisticated technologies and equipment, the development of the airplane being the prime example. The difficulty to use this 'high-tech' equipment created an interest in human capability, as some people were clearly better than others in certain specific tasks. Psychologists started to engage with this problem and devised numerous aptitude and capability tests—the intelligence (IQ) test being the most well known. These tests were to find the right person for specifically outlined jobs or task. During World War I, for example, psychological test centres were erected all around the world to select pilots, sound detector operators, and other highly specialised jobs (Meister, 1999). Note that people were tested, selected, and trained for certain pre-defined existing jobs, but still no efforts were made to adjust the task or equipment to best suit human capabilities (Meister, 1999; Roscoe, 1997).

At the time of Word War I, we also saw the rise of Taylorism, or Scientific Management, which was one of the first attempts to apply scientific principles to the design of processes to increase the output the workplace (i.e. factories). Taylor (1967) used formal methods of data collection and statistical analysis—ones that are not far removed from those that human factors professionals use today (Meister, 1999)—to design the optimal manner of organising work.

> Perhaps the most prominent single element in modern scientific management is the task idea. The work of every workman is fully planned out by the management at least one day in advance, and each man receives in most cases complete written instructions, describing in detail the task which he is to accomplish, as well as the means to be used in doing the work. And the work planned in advance in this way constitutes a task which is to be solved, as explained above, not by the workman alone, but in almost all cases by the joint effort of the workman and the management. This task specifies not only what is

> to be done but how it is to be done and the exact time allowed for doing it. … The task is always so regulated that the man who is well suited to his job will thrive while working at this rate during a long term of years and grow happier and more prosperous, instead of being overworked. Scientific management consists very largely in preparing for and carrying out these tasks. (Taylor, 1967, p. 39)

By breaking up larger tasks into increasingly smaller ones, and subsequently engineering their optimal manner, timing, and sequence, Taylor aimed to increase the productivity and wellbeing[3] of workers by improving the efficiency of workflows in the workplace[4]. As such, scientific management can be seen as a first attempt to modify the system, or the processes, in which people are working, rather than chiefly focusing on the human operator. This point is further illustrated by Taylor: “Under the old type of management success depends almost entirely upon getting the "initiative" of the workmen, and it is indeed a rare case in which this initiative is really attained” (1967, p. 36).

World War II, again, brought an entirely new level of complexity to socio-technological systems. Mostly due to the enormous amount of people involved in World War II, it became increasingly insensible to keep relying on the strategy of selecting people to fit a pre-defined task (Meister, 1999). The novel idea was to adjust the task or tools to match the human, that is, to design tasks and equipment that takes into consideration the limits of human capabilities. This idea came out of practical examples in which such adaptation of equipment lead to increased human-technology performance (and thus safety).

One of such examples—often told as an amusing anecdote in introductory human factors courses—is the story of psychologist Lt. Alphonse Chapanis, who is said to be (one of) the first to apply psychological principles to help solve a design problem in airplanes that led to many ‘pilot errors’ during the war. The particular problem that Chapanis faced was that the pilots of a particular aircraft type, the P-47, B-17 and B-25,

---

[3] Taylor’s words, make workers ‘grow happier and more prosperous’, are thus in line with the explicit aims of human factors (see the IEA definition at the beginning of this chapter).

[4] Whether this initial form of ‘proceduralisation’ actually helps to increase the (mental) wellbeing of workers is a second discussion. One I briefly touch upon in my book chapter, ‘Human factors and the ethics of explaining failure’, of which a first draft is attached to this thesis as an appendix.

after landing frequently retracted the gear instead of the flaps. This, logically, had the necessary operational and financial consequences. The interesting observation that Chapanis made was that this same problem hardly ever occurred with C-47 pilots. Upon closer inspection of the cockpit of the problematic group of airplanes (P-47, B-17, B-25), he noticed that the gear and flap controls could easily be confused, as they were nearly identical levers. Moreover, they were located right next to each other. The corresponding controls in the C-47 cockpit, however, were not adjacent and thus not easily confusable. Chapanis realised that the so-called 'pilot errors', the gear-up-after-landing problems, were therefore actually cockpit-design errors, and that coding the shapes and modes of operation of the control levers should easily prevent future accidents. When a small rubber wheel and a small wedge-shaped end were fixed to respectively the landing gear lever and the flap lever, this type of 'pilot error' almost completely disappeared (Chapanis, 1999; Roscoe, 1997).

This World War II insight, that we have to design equipment and technology to support human capabilities in order to avoid the negative effects of human limitations, effectively constitutes the origin of the scientific discipline of human factors (Batteau, 2001; Dekker, 2005; Meister, 1999). After the war the scientific discipline of human factors really took off, as the post-war climate (the cold war period) was extremely favourable for government-supported research. Besides the availability of research funding, the rapid technological developments that the World War II had initiated continued, raising human performance questions that demanded research explanations and design solutions (Meister, 1999). Consequently, many experimental psychology laboratories were established. One of such, the Psychology Branch of the Aero Medical Laboratory, was lead by an experimental psychologist, Paul Fitts, who was, together with Alphonse Chapanis, one of the founding fathers of the human factors discipline. His book 'Psychological Research on Equipment Design' (see Fitts, 1947) was the first major publication on human factors engineering (Meister, 1999).

A seminal study that placed human factors on the map as a science concerned with the interactions between people and technology, was by Fitts and Jones (1947). Besides marking the beginning of contemporary human factors research, it also introduced a new approach to (aviation) accident investigation (Coury et al., 2010). They introduced their new idea for dealing with 'pilot error' in the following manner:

> Aircraft accidents usually are classified as due to pilot error . . . It has been customary to assume that prevention of accidents due to materiel failure or poor maintenance is the responsibility of engineering personnel and that accidents due to errors of pilots or supervisory personnel are the responsibility of those in charge of selection, training, and operations. The present study was undertaken from a different point of view; it proceeded on the assumption that a great many accidents result directly from the manner in which equipment is designed and where it is placed in the cockpit, and therefore can be eliminated by attention to human requirements in the design of equipment. (Fitts & Jones, 1947, pp. 335-336)

Entirely in line with Chapanis (1999), Fitts and Jones (1947) thus assumed that human performance is systematically connected to features of people's tools and tasks. As such, the goal of human factors is to design equipment in a manner that supports human capabilities in order to overcome human performance limitations (and 'human error').

Fitts and Jones were experimental psychologists who adapted their laboratory techniques to applied problems, and in doing so created an important new way of analysing accidents and human performance issues. They introduced (experimental) psychology's empirical-experimental methodology to the study of real-world complex social-technical problems. Their work displays a clear shift from the focus on training and selection, to the inclusion of human factors in the design of technology (human-centred design). In doing so they set the agenda for the scientific discipline of human factors. The foundational insight that human performance is systematically connected to features of people's tools and tasks was also to provide a new way of thinking about safety, and accident investigation in particular. Now it was no longer sufficient to conclude an accident investigation by identifying the human (pilot) error as the cause, as the design of the system and its contribution to the error had to be included as well (Coury et al., 2010).

1.4 Accidents models and the human contribution to accidents

Accident models provide a framework for explaining how accidents happen and, consequently, how we should try to prevent them (Leveson, 2002). Heinrich (1941) set the stage with his domino model and its search for manners to intervene in the chain of events to prevent the eventual accident. However, in 1970s and 1980s safety scientist had to revise Heinrich's linear approach to afford for more complex and social realms in their models. This was partly the result of a number of large (industrial) accidents that displayed the limits of our ability to control our modern socio-technological systems and turned safety into a topic of social interest (Coury et al., 2010). Accidents such as the Tenerife Airport disaster (1977), the Three Mile Island nuclear incident (1979), the Bhopal chemical accident (1984), the space shuttle Challenger blow-up (1986), the Chernobyl nuclear accident (1986), the capsizing of the MS Herald of Free Enterprise (1987), London's King's Cross station fire (1987), the Piper Alpha rig blowout (1988), and the Clapham Junction train crash (1988) dramatically showed the limits of our technical and overly linear models for understanding failures in modern, highly automated socio-technological systems.

Take, for example, the Three Mile Island (TMI) incident. Although there were no injuries, the (near) disaster was a watershed event for nuclear power and safety in general. On March 28, 1979, maintenance workers at the TMI nuclear power plant accidentally blocked the coolant water to the reactor core, resulting in the core's temperature rapid rise. As designed and trained for, the plant automation and operators initiated an emergency shut-down, turning on emergency water pumps to prevent the core from further heating and eventually melting. However, as their displays did not show this, the control room operators did not realise that the valves for the emergency water flows were shut, preventing emergency coolant water to reach the core. As the coolant water dropped further, operators struggled to make sense of the conflicting messages and alarms they received in the control room. By the time they had finally restored the coolant water flow, the reactor vessel had filled up with hydrogen gas, where, together with the high temperature and pressure, it created a dramatic threat of a blowout, which could result in the spread of radioactive material all over the plant's wide surroundings. Consequently, more than a hundred thousand people were evacuated, as it took several days to neutralise the hydrogen gas and prevent a disaster (Coury et al., 2010; Perrow, 1981, 1999).

These major disasters showed that highly automated systems with multiple layers of (technical) redundancy—system that were assumed to be fail-safe—were still susceptible to failure. Accidents as TMI, that displayed our inability to control new and more complex technologies, drastically changed the disciplines of human factors and safety science. Public demand for explanations of these dramatic events played a central role, resulting in vast amounts of research into the social and organisational contribution to accidents. In the wake of the Tenerife airport disaster, for example, in "the 1970's, an new theme became popular in accident investigations: that of small group interactions. More specifically, it became legitimate to consider, for example, 'lack of crew resource management' as [sic] a valid contributor to mishap" (Hummerdal, Wilhemsson, & Dekker, 2013, p. 409). Similarly, one of the key findings into the causes of the Chernobyl disaster was that production pressures had lead to 'culture' in which inadequate attention was given to safety (International Nuclear Safety Advisory Group, 1986) (see chapter 8 for the analysis of 'safety culture' as a discursive object).

The catastrophic events in the late 1970s and 1980s thus showed the need for a more complex understanding of accidents in our modern socio-technological systems, moving away from earlier accident models that focus chiefly on technical reliability, towards models that allow for an understanding that accounted for the human and organisational contribution to accidents. In line with the human factors principle outlined by Chapanis (1999) and Fitts and Jones (1947) to see human performance as systematically connected to people's tools and tasks, Leveson explains about accident models:

> … devising more effective accident models will require shifting the emphasis in explaining the role of humans in accidents from error (deviations from normative procedures) to focus on the mechanisms and factors that shape human behavior, i.e., the performance-shaping mechanisms and context in which human actions take place and decisions are made. (Leveson, 2004, p. 246)

This need for a more complex understanding of the role that humans play in the causation of accidents is further illustrated by Le Coze, who explains that some of "the most popular models, still widely influential today, were established in the end of the 1980s … These prevalent models of the past 20 years are in fact the products of works by psychologists, social psychologists and cognitive ergonomists." (Le Coze, 2013, p. 202). With this focus on the social contribution to accidents, the disciplines of human factors and safety science thus converge even further in their orientation.

Many of these new of accident models are built on the vast body of work of Jens Rasmussen (Rasmussen, 1983, 1990a, 2000). Probably his most renowned contribution to human factors and safety science is his framework for categorising human performance as skill-, rule-, or knowledge-based (Rasmussen, 1983). He linked these performance categories to corresponding error prone situations, design solutions, and other remedies. Rasmussen (1983) also explicitly stated that humans are teleological by nature, that is, they are goal-oriented beings who actively select goals and information. This simple insight meant a break from seeing humans as the random creators of dangers, as well as the ones who needed protection against hazards, towards an more appreciative role for the human actively creating safety. With this insight, Rasmussen paved the way for safety thinking that recognises that humans are not the greatest threat to safety, but rather that they work in an environment of uncertainty and complexity and that they thus need to conciliate between multiple completing goals in order to forestall failure and accidents. In running ahead of my narrative, Rasmussen provided the foundation for safety thinking that sees people are the greatest resource, not the greatest threat, to the safety of our systems (e.g. Dekker, 2002a; Woods, Sarter, Cook, & Johannesen, 1994).

James Reason (1990, 1997) developed Rasmussen's (1983) behavioural performance levels to propose a number of error types: skill-based slips and lapses, rule-based mistakes, and knowledge-based mistakes. Reason (1990) further extended his elaborate error taxonomy by distinguishing between 'active errors' and 'latent failures,' the latter he later (in his 1997 book) called 'latent conditions'. The effects of active errors are felt almost immediately, however, latent errors or conditions "may lie dormant within the system for a long time, only becoming evident when they combine with other factors to breach the system's defences" (Reason, 1990, p. 173). A latent condition, which Reason (1990, p. 197) likened to "resident pathogens" in the body, can thus be a design defect in hardware or software, but also people and social systems have inherent flaws that only manifest themselves when they combine with other factors to break through the system's defences.

This idea of latent failures brought Reason to propose the 'Swiss Cheese' accident model, which has made a large impact on accident investigation and safety management all over the world. For Reason (1990) the accidents in the late 1970s and 1980 made it clear that the performance of those at the sharp end—the front line operators—was shaped by latent conditions and upstream organisational factors. Besides drawing on Heinrich's (1941) notion of a 'hazard', a source of 'energy' that needs to be contained, the Swiss Cheese model further borrows from Heinrich's Domino model in seeing an accident as a sequences of events: a sequence of safety barriers that get breached. Reason (1997) explains that systems have many barriers, or layers of defence—at the design, organisational, management, or operator level—to protect against hazards, but that these 'defences in depth' can never be perfect. These flaws in the defences are the holes in the slices of the Swiss cheese (the slices being the various defences in depth). Reason completed his Swiss Cheese metaphor by depicting an arrow passing through successive holes in the slices of cheese to show the progression of an accident as a series of latent and active failures penetrating the defences in depth (Reason, 1990, 1997).

The Swiss Chees model convincingly conveys the idea that no single failure, human or technical, is sufficient to cause an accident. Rather, it involves the combination of several contributing factors arising from various levels of the organisation or system. As such, just like for Chapanis (1999) and Fitts and Jones (1947), human error is no longer simply a category in a causal taxonomy, but instead reflects a more complex interaction between people and their tools and tasks. Simply put, Reason (1997) suggests that, beyond their tools and tasks, the performance of those at the sharp (operational) end is shaped by latent conditions and upstream organisational factors.

1.5 Systems thinking in human factors and safety science

Even though Reason's (1997) Swiss cheese model tries to include multiple interactions, these are always represented as a chronological chain of events, as an 'arrow' flying through the holes in the cheese. The linearity of the model is further displayed by its derivatives, such as root cause analysis (RCA) or other types of fault trees, which try to trace back the sequence of contributing events to arrive at the triggering event, the (root) cause(s) of the accident. The linearity of Reason's model makes it susceptible to some of the same critique as Heinrich's (1941) domino model, which basically comes down to the argument that this approach to accident causation remains committed to an overly simplistic and reductionist idea of looking for broken components and errors (active or latent) in understanding accidents (e.g. Hollnagel, 2014b; Woods, Dekker, Cook, & Sarter, 2010).

During the twentieth century a number of theorist, across various disciplines, became increasingly suspicious of the reductionist method and started to advocate a more ‘holistic’ or ‘systems’ approach to the various problems they encountered. As opposed to Descartes's method of analytic reductionism, which dictates that the whole can be understood by analysing the properties of its parts, these more holistic approaches see systems as an integrated whole whose properties ‘emerge’ from the parts. Characteristics or phenomena of the system can therefore not be reduced to the functioning of its components (Capra, 1982; Capra, 1996). With the rise of new multi-disciplinary ideas and research agendas, such as cybernetics (Ashby, 1956) and general systems theory (von Bertalanffy, 1969), systems thinkers also turned to safety and accident modelling.

In saying that, “[s]afety is an emergent property of systems and not of their components” (Woods & Cook, 2002, p. 140), systems thinkers are not satisfied with models that explain accidents as a sequence of events in which the events almost always are some type of component failure or human error. This type of linear causality is too simple for the increasing complexity of our rapidly changing socio-technological systems (Leveson, 2004). Systems thinkers argue that the causal factors that are identified in such event-based accident models depend on the events and conditions that are initially considered, which are often only the proximate events that precede the accident. The choice of events to include in the analysis is thus necessarily subjective, and the triggering event that is often labelled the ‘initiating event’ is an arbitrary choice, as preceding events or conditions for that initial event can always be found (Dekker, 2006). Moreover, systems safety theorists are unwilling to point to one particular event as a ‘cause’, saying that if the ‘triggering event’ had not happened another one would have started the ‘chain of events’. To go back to Rasmussen, who, as explained in the previous section, can be seen as the father of this kind of systems thinking in safety:

> The important issue is that the stage for an accidental course of events very likely is prepared through time by the normal efforts of many actors in their respective daily work context, responding to the standing request to be cost-effective. Ultimately, a quite normal variation in somebody’s behaviour can then release an accident. Had this particular ‘root cause’ been avoided by some additional safety measure, the accident would very likely be released by another cause at another point in time. In other words, an explanation of the accident in

> terms of events, acts, and errors is not very useful for design of improved systems. (Rasmussen, 1997, p. 190)

In line with Rasmussen's argument, Charles Perrow (1999, first published in 1984) introduced the idea of a 'system accident' for explaining accidents that are not the result of an initial 'triggering event' or component failure, but rather as the inevitable result of the complex interactions among various tightly coupled processes or components. Perrow (1999) explains that at Three Mile Island, it was not the accidental closure of the coolant water valve that caused the accident, but that it were the interactions between (poor) maintenance procedures, (poor) design of the control room and panels, and the operators (lack of) understanding of the automation (closing the feed-water valves) that brought Three Mile Island so close to a major disaster. Perrow argues that in interactively complex and tightly coupled systems accidents occur as the normal by-product of production, hence the name of his theory, 'Normal Accident Theory' (NAT). When a system is interactively complex, independent events (failures) can interact in unforeseeable ways, which are incomprehensible for the designers of and operators in the system. When a system is also tightly coupled, meaning that its components are highly interdependent, these unforeseen interactions can rapidly spread trough the system and have devastating consequences. Perrow thus argues that (normal) accidents emerge from the characteristics of the system itself, and they cannot be anticipated, prevented or trained against (Leveson, Dulac, Marais, & Carroll, 2009; Perrow, 1999). As such, Perrow and other systems theorists argue that Reason's (1997) 'defences in depth' may have outlived their usefulness, as such barriers only protect against simple, foreseeable failures and may even be detrimental to safety, in that they might increase the complexity of the system that they are trying to protect (Dekker, S. W. A., 2011; Perrow, 1981, 1999).

As a reaction to the negative perspective that Perrow offers—saying that accidents are 'inevitable' in our modern complex socio-technological systems—a group of Berkeley safety researchers, Gene Rochlin, Todd La Porte, and Karlene Roberts (1987), provided a counter example, that of aircraft carrier flight operations. Their argument was that for such a tightly coupled and complex organisation this system was not having nearly as much accidents as Normal Accident Theory would predict: "there is a small group of organizations in American society that appears to succeed under trying circumstances, performing daily a number of highly complex technical tasks in which they cannot afford to ‘fail’” (Rochlin et al., 1987, p. 76). Rochelin et al. (1987) termed this group ‘High Reliability Organizations’ (HROs).

High Reliability Theory (HRT) identifies four organisational characteristics that promised to decrease organisational accidents (Rochlin et al., 1987). First, the organisation, and especially its management, has to acknowledge and promote safety and performance as the primary goals of the organisation (we will see this feature resurface in the discussion of safety culture). The second point HRT makes is that operations have to be simultaneously centralised and decentralised: HROs have to be centralised in that there has to be a clear chain of command, however, HROs also give a considerable amount of authority to low-level personnel, as “[h]ierarchical rank defers to the technical expertise often held by those of lower formal rank.” (LaPorte & Consolini, 1991, p. 32). The HRO researchers observed that on an aircraft carrier even the lowest-level seaman could abort a landing, as this local authority (decentralisation) is necessary when decisions must be made too quickly to go up the chain of command (Rochlin et al., 1987). A third characteristic of HROs is that they use ‘sophisticated trial-and-error’ as a means of organisational learning (LaPorte & Consolini, 1991). In an organisation where the cost of failure is to high, they use imagination, simulations, exercises, and story telling, as an alternative to trial-and-error learning. Finally, HROs seem to rely on multiple forms of redundancy, which is “the ability to provide for the execution of a task in the primary unit fails or falters” (Rochlin et al., 1987, p. 84). This operational redundancy, either technical or social, can be achieved through duplication, two units performing the same task, or overlap, two units with functional areas in common (LaPorte & Consolini, 1991; Leveson et al., 2009; Rochlin et al., 1987).

Since the Berkeley group first presented their ideas, there has been a great deal of critique, both from outside and inside the HRT group (see Rochlin, 1999). Besides the critique that the organisations that HRT studied were not really ‘tightly coupled systems’ (LaPorte & Consolini, 1991; Leveson et al., 2009), the main problem that many systems thinkers explain to have with HRT is that it seems to equate reliability with safety. Reliability is about whether a component, human or technical, lives up to pre-specified performance criteria (Dekker, S. W. A., 2011). Therefore, if an operator is not following specified procedures, he can be said to be operating unreliably. However, in some cases it might be detrimental to safety to comply with prescribed procedures—see, for example, the Swiss Air 111 accident (Transportation Safety Board of Canada, 2003). Systems thinkers argue that safety and reliability are different qualities that should not be confused. “This is one of the basic distinctions between safety and reliability: Safety can only be determined … in the context of the whole. Therefore, it is not possible to take a single system component in isolation and assess its safety” (Leveson, 2004, p. 249).

The idea that accidents often result from the interactions of normally (reliably) functioning components is the central notion in the most recent group of accident models and human factors thinking (e.g. Dekker, S. W. A., 2011; Snook, 2000; Vaughan, 1996). These models or theories react to the afore discussed explanations for accidents that keep relying on the reductionist idea of explaining failure at the systems level—an accident—by pointing at failure at the component level—most prominently human error. To an extent, even Perrow’s (1999) Normal Accident Theory suffers from this critique, as he also relies on the (interaction of the) failures of components to explain why accidents occur. Systems thinkers are keen to point out that in complex systems accidents often emerge from the interaction among normally functioning components, and that we thus need not to look for failures at the ‘micro’ level to understand ‘macro’ events as accidents: “systemic accident models do not rely on a component breaking or a human erring. In fact, they do not have to rely on anything "going wrong", or anything being out of the ordinary” (Dekker, 2006, p. 92)

Instead of seeing accidents as the result of component failures, systemic accident models see them as a the ‘by-product’ of the normal functioning of the system (Dekker, 2006; Dekker, S. W. A., 2011). Pidgeon and O’Leary, for example, in advocating for Turners ‘man-made disaster theory’ explain that accidents may emerge out of normal processes: “The simple message of man-made disasters theory is that, despite the best intentions of all involved, the objective of safely operating technological systems could be subverted by some very familiar and ‘normal’ processes of organizational life” (Pidgeon & O'Leary, 2000, p. 16). Similarly, Vaughan, Snook, and Dekker—in drastically simplifying their respective theories ‘Normalisation of Deviance’ (Vaughan, 1996), ‘Practical Drift’ (Snook, 2000), and ‘Drift into Failure’ (Dekker, S. W. A., 2011)—argue that organisational pressures (or culture) for production and efficiency slowly but systematically shape the decision making and performance of those in the system. As these pressures slowly influence people’s trade-offs, it incrementally ‘normalises’ what was previously seen as unacceptable and dangerous. As such, according these models, ‘normal people doing normal work’, and parts functioning within their specified limitations, can still create tragic outcomes (Dekker, 2004).

In looking at the human contribution to accidents, to explain failure we should not look for where people went wrong, but instead, we should try to understand how people's assessments, decisions, and actions made sense given the circumstances that surrounded them. As earlier expressed by Rasmussen, systems safety thinking does not see our complex socio-technological systems as inherently safe: our modern socio-technological systems are pervaded by uncertainty, all sorts of constraints, contradictions, and goal-conflicts (Dekker, 2001). “Contradictory goals are the rule, not the exception, in complex situations” (Dörner, 1996, p. 65). As such, it comes down to the people in the systems to conciliate these multiple competing goals, principally to balance production and safety (Hollnagel, 2009). “People are doing reasonable things given their point of view and focus of attention; their knowledge of the situation; their objectives and the objectives of the larger organization they work for” (Dekker, 2006, p. 13). This assumption, labelled the ‘local rationality principle’ (Dekker, 2002b; Woods et al., 1994), is not only a central premise of the latest thinking in safety science, it is also entirely in line with the ideas for a human factors science as outlined by Chapanis (1999) and Fitts and Jones (1947).

Chapanis (1999) and Fitts and Jones (1947) argued that we should see human performance as systematically connected to people’s tools and tasks, and that we thus need to design equipment in a manner that supports human capabilities and minimises possibilities for ‘error’. The latest safety thinking, which argues that organisational pressures for production and efficiency slowly but systematically shape the decision-making and performance of those in the system, extends this human factors thinking to the organisational—or systems—level: In addition to people’s tools and tasks, human performance and behaviour is connected to the operational and organisational environments people work in. Just as the safety of the system cannot be understood by looking for failing components, so can human performance only be understood by seeing it as embedded in the larger system in which it takes place. As such, human factors and safety science build on a systemic worldview.

1.6 Conclusion

Anno 2015, accidents are no longer understood as the result of (a linear succession of) failing components, whether human or technological. Rather they are seen phenomena that can emerge out of the normal functioning of organisations—systems. Similarly, human performance should be understood as part of the larger operational and organisational system in which people are configured. Beyond their shared concern with the safe and efficient functioning of our modern socio-technological systems—with a central concern for the wellbeing of the people in and dependent on these systems—human factors and safety science also share their historical trajectory towards systems thinking. As powerfully expressed by Woods and Cook (2002, p. 140), "safety is created and sometimes broken in systems, not individuals. The issue is finding systemic vulnerabilities, not flawed individuals."

This is how far we have come with human factors and safety science as systems sciences in the 21st century. Almost 70 years after the seminal work of Chapanis (1999) and Fitts and Jones (1947) (even longer since Heinrich (1941)), there are only a few researchers or practitioners in human factors that do not think of it being a systems discipline (Wilson, 2014). In 2010 the International Ergonomics Association (IEA) put together a small working group to address the questions of how human factors should move into the future as a discipline and profession (Dul et al., 2012). They concluded that the best way for the discipline to distinguish and strengthen itself is through its system orientation. "The focus of HFE is to jointly improve performance and well-being by designing the integrative whole better, and by integrating the human into the system better. This is done by fitting the environment to the human" (Dul et al., 2012, p. 378). Or as Wilson (2014, p. 56) put it: "It is its very systems perspective and holistic nature that provides the strength of ergonomics". He elaborates by explaining, "to cover all aspects of people' s interaction with their environments and the interconnections between these interactions, is what allows it to define itself as a unique discipline" (2014, p. 56).

Similarly, safety science in the 21st century can be characterised by systems thinking—"It has a holistic character" (Le Coze et al., 2014, p. 1). In line with the narrative of this chapter, Hollnagel (2014a; 2014c) explains that as our socio-technological systems have become increasingly complex throughout the 20th century, our older (reductionist) accident models have become inept. "Simple linear accident models, represented by Heinrich's domino model, are well-suited to situations that resemble what work was like in the 1920s and 1930s, but not to the 1970s and beyond" (Hollnagel, 2014c, p. 23). Similarly, 'composite linear models', such as Reason's Swiss cheese model, were well suited for understanding work was like in the 1970s and 1980s. However, "[m]odels and methods which require that systems are linear with resultant outcomes cannot and should not be used for non-linear systems where outcomes are emergent rather than resultant" (Hollnagel, 2014c, p. 23). The large industrial accidents in the 1970s and 1980s have shown the need for models that take into account the complexity of our systems and the 'locally rational' behaviour of the people in it (Dekker, 2010b; Dekker, S. W. A., 2011; Vaughan, 1996). "If we tinker only gingerly with the final, marginal technical minutiae at the sharp end, all of those systematic influences will collect again and again to shape what any [operator] will see as the most rational course of action" (Dekker, 2010b, p. 148). As such, just as human factors is now generally considered to be based on a systems approach, in safety science, "[t]here is now a widely adopted statement that 'safety is an emergent property'" (Le Coze et al., 2014, p. 4).

The larger movement of systems thinking has thus manifested in both human factors and safety science. It is in this light of systems thinking that we should see the endeavours—the explanations, classifications, measures, design recommendations, interventions, etc.— that human factors and safety science propose. This thesis asks what insights Foucault's archaeological analysis of the formation of objects can provide the disciplines of human factors and safety science, and if these insights can help to critically reflect on some of the discursive practices that these disciplines promote. Together with the explicit goals of pursuing system efficiency and the wellbeing of people in the system, it is thus in the light of this systems thinking that some prominent human factors and safety science objects need to be assessed.

As prominent manifestations of both the human factors and safety science discourse, this thesis will provide archaeologically inspired studies of the scientific objects of situation awareness, safety science, and resilience (engineering). I will not introduce these objects here to avoid unnecessary duplication, as they will be extensively described in the analytical chapters, presented in Parts III and IV. Suffice it to say that, to a greater or lesser extent, these three objects all claim to be drawing on a systems approach with the intent to improve the efficiency of the system as well as increase the wellbeing of the people in it.

Chapter 2.
"Loss of situation awareness" by medical staff: Reflecting on the moral and legal status of a psychological concept

Statement of contribution to co-authored published paper

This chapter includes a co-authored paper. The copyright of this article has been transferred to Thomson Reuters, however, authors have the right to use their article for non commercial purposes, such as the inclusion in a dissertation.

The bibliographic details of the co-authored paper, including all authors, are:

Breakey, H., Winsen, R. van, & Dekker, S. W. A. (2014). 'Loss of situation awareness' by medical staff: Reflecting on the moral and legal status of a psychological concept. Accepted by the *Journal of Law and Medicine*.

As a co-author of this paper I have had a leading role in pursuing this argument; I have reviewed the literature on SA and criminal negligence; I have written the first drafts of this paper; I have engaged the other authors; I have critically revised numerous versions of the paper; I have contributed to the structure and outline of the final draft of the paper; I have written a major part of the text of the final, accepted, paper.

(Signed)________________ (Date) 9-2-15
Roel van Winsen

(Countersigned) ________________ (Date) 9.2.15
Hugh Breakey (Associate supervisor and corresponding author)

(Countersigned) ________________ (Date) 9.2.15
Sidney W. A. Dekker (Principal supervisor and co-author)

Prescript to paper

This chapter presents an accepted peer-reviewed journal article, *"'Loss of situation awareness' by medical staff: Reflecting on the moral and legal status of a psychological concept"*, which argues that invoking the object of situation awareness (SA) becomes problematic when its discursive use broadens to include professional norms for behaviour and is used to judge performance in questions of criminal negligence and culpability.

SA, which commonly refers to "knowing is going on" (Endsley, 1995b, p. 36)[5], was once introduced with the purpose of optimising human-machine interface design and understanding human cognition in our increasingly opaque and complex socio-technological systems. The use of SA in the design of human-centred interfaces is entirely consistent with the goals and principles of human factors and safety science (as outlined in the previous chapter). However, after the object was introduced as a specifically defined object of the human factors and safety science discourse, a number of self-reinforcing discursive practices started delimiting, defining, and legitimizing it further. During the 1990s SA rapidly gained in popularity both within and outside human factors, and it started to appear in other discourses where its discursive use was significantly broadened.

Shortly after its conception, SA—or 'loss of SA'—started to appear in (aviation) accident investigation reports where it was attributed with causal power to explain human behaviour resulting in accidents. In the 1990s SA emerged, for example, in various accident reports surrounding the particularly vexing aviation problem of 'controlled flight into terrain' (CFIT). This sudden prevalence was probably due to the introduction of new technology—the enhanced ground proximity warning system, which aims to support the flight crew in their awareness of the terrain they are flying over. A joint project by the International Civil Aviation Organisation (ICAO), Flight Safety Foundation, and Federal Aviation Administration stated in 1995: "More than two-thirds of all CFIT accidents are the result of altitude error or lack of vertical situational awareness" (p. 4.C.6). These numbers were corroborated by a study conducted by the Netherlands National Aerospace Laboratory (NLR) that looked into worldwide CFIT accidents

[5] In later chapters SA will be extensively and redundantly described, so for now I choose not to add to the duplication and keep my description of SA to a minimum.

between 1988 and 1994. Additionally the NLR stated that SA errors were among "the dominant crew-error types" and that in the special case of approach phase accidents, "virtually all the accidents involved a situational awareness error" (Khatwa & Roelen, 1996, p. 16). Similarly, the Australian Transportation Bureau (Australian Transportation Safety Bureau, 2007, p. 50) called SA the "primary cause" in their study of CFIT accidents in Australia between 1996 and 2005.

A more recent example of how SA is used as a causal construct is the investigation report by the Brazilian national aeronautical accident investigation board (CENIPA, 2008) into the mid-air collision of a Gol Linhas Aéreas Boeing 737 and a Embraer Legacy business jet over Brazil's Amazonia in 2006. Reference to SA, in particular notions as 'loss of SA', 'poor SA', 'low SA', and 'inadequate SA', occur frequently throughout the document (SA is referred to 45 times) and carries a substantial part of the causal explanation for why both the pilots and the controllers behaved the way they did. The report concluded that: "The low situational awareness of the pilots (airmanship) was a relevant factor for the occurrence of the accident" (p. 259). However, the causal power attributed to SA in this report is probably most strongly exemplified by the following paragraph: "The transponder switchoff was not perceived by the crew, due to the reduction of the situational awareness... The lack of situational awareness also contributed to the crew's not realizing that they had a communication problem with the ATC" (CENIPA, 2008, p. 159).

The discursive use of SA as a normative, causal construct is just one way in which the object has discursively 'broadened' (a more elaborate discussion about this specific problematic use of SA as a causal construct is provided by Billings, 1995; Dekker & Hollnagel, 2004; Flach, 1995b; Moray & Inagaki, 2000). A related, equally problematic, manner is presented in the journal article that follows. It presents the discursive use of SA in questions of moral culpability and criminal liability.

In thinking of an audience for our argument, the co-authors and I decided to (first) pitch it specifically to the medical domain. As such, instead of providing a general discussion about the problematic use of SA as an object[6], the article provides (just) one—

[6] Initially, this article was larger in scope—it provided a cross-disciplinary description of the changing discursive use of SA and the manner in which the object enters the juridical discourse. The article was also based on an explicit Foucauldian archaeological framework. However, to enhance the chances of

powerful—instance of the problematic discursive use of the object SA. Within the scope of the thesis, the main purpose of this article is to show that some of the objects that are proposed with the best of intentions—to help design human-machine interfaces and explain human cognition in complex dynamic environments—can function in ways that are not only inappropriate (as we argue in the article) but also run counter to the explicit aims of these objects and the larger discourses (i.e. disciplines) that they are part of.

publication, we significantly reduced the length of the article and excluded the Foucauldian framework. At a later stage, I will pursue the publication of the larger, cross-disciplinary, argument.

“Loss of situation awareness” by medical staff: Reflecting on the moral and legal status of a psychological concept

Abstract

This paper examines the emergence of “accurate situation awareness (SA)” as a legal and moral standard for judging professional negligence in medicine. It argues that SA constitutes a status, an outcome resulting from the confluence of a wide array of factors, some originating inside and others outside the agent. SA does not connote an action, a practice, a role, a task, a virtue, or a disposition—the familiar objects of moral and legal appraisal. The argument contends that invoking SA becomes problematic when its use broadens to include professional or legally appraisable norms for behaviour, which expect a certain state of awareness from practitioners.

2.1 Introduction

“Accurate situation awareness (SA) of medical staff is integral for providing optimal performance during the treatment of patients” (Schulz, Endsley, Kochs, Gelb, & Wagner, 2013, p. 729). Thus proposes the recent issue of an anesthesiology journal. In this paper, we consider the emergence of “accurate SA” as a legal and moral standard for judging professional negligence. SA, though there are a number of characterizations, refers broadly to a practitioner’s knowledge and understanding of a situation and its evolution, informed by her or his perception and attention. In other words, “accurate SA” means the agent holds correct beliefs about the relevant elements of the situation, while “mistaken SA” or “poor SA” connotes the agent holds false or incomplete beliefs about those elements. We take the notion of ‘false belief’ in the strict positivist manner of most SA research and use of SA in investigation reports, meaning that a person’s internal belief, their representation of the world, does not correspond to the external material world (Dekker, 2005).

We contend that SA should play no role assigning blame or, more specifically, in determinations of whether a professional has breached their duty of care. In ordinary understandings of the duty of care, whether or not there is a breach of the duty depends on several factors, such as foreseeability of harm, magnitude of the risk, the practicality of the precautions taken, and of reasonableness (Corcoran, 2000). Ultimately, the question focuses on whether a hypothetical reasonable person would have acted in a similar manner under the same circumstances. "The test is not whether the defendant acted reasonably, but whether a reasonable person placed in the position of the defendant would have acted as the defendant did" (Michaelides-Mateou & Mateou, 2010, p. 11). Our question in this paper is whether SA should be injected into appraisals of a breach of duty of care. In other words, we question whether appeals to "accurate SA", "poor SA" or "mistaken SA" should play a role in determining moral and legal blameworthiness.

Accident inquiries outside of medicine have long used SA in their causal reasoning, connecting deficient or inaccurate or lost SA with a bad outcome. The investigation into a Brazilian mid-air collision, for instance, concluded that "[The pilots] did not monitor the functioning of the transponder, and maintained a poor situational awareness, well below the recommended standards…" (CENIPA, 2008, p. 253).

In the usual course of affairs, professionals are judged according to the standards of their particular profession. A pilot, air traffic controller, doctor, engineer, or social worker must thus be judged for his or her reasonableness, respectively, through the eyes of the reasonable pilot, air traffic controller, doctor, engineer, or social worker. This principle has become known as the Bolam Test:

> Where you get a situation which involves the use of some special skill or competence, then the test as to whether there has been negligence or not is not the test of the man on top of the Clapham Omnibus because he has not got any special skill. A man need not to possess the highest of expert skill; it is well established in law that it is sufficient if he exercises the ordinary skill of an ordinary man exercising that particular art (Justice McNair, 1957, in Bolam v Friern Hospital Management Committee, as read in Michaelides-Mateou & Mateou, 2010, p. 12).

The Bolam test says that a professional is not required to have an extraordinary degree of expertise or knowledge, but a degree of competence common to that of the "ordinary skilled man" in that profession (Corcoran, 2000).

In both accident inquiries and legal prosecutions, SA has been employed as one element of this professional competence. The coroner who investigated a friendly fire incident which killed three British soldiers in Afghanistan in 2007, rendered the verdict that the crew of an American fighter jet had lost "situational awareness" and were looking at the wrong village when they dropped the bomb (Bruxelles, 2010). In a case in Canada, a maritime officer was criminally convicted for his role in a ferry accident that killed two passengers. The argument for his dereliction turned on the ordinary skilled standard—that of the reasonable maritime officer—which was held to dictate that, "[m]aintaining situational awareness at all times and in all circumstances is key to proper navigation" (Supreme Court of British Columbia, 2013). An 'ordinary skilled' doctor, by comparison, could be required to have accurate SA, on the basis that it is "integral for providing optimal performance." In order to avoid liability for negligence, doctors would therefore have to be able to show or argue that they had "accurate SA."

In hindsight (Hugh & Dekker, 2009) it is always easy to blame a "loss of situation awareness". As Moray and Inagaki warned in 2000 (in a quote where "investigators" might easily be replaced by "judges"):

> 'situation awareness'…is poorly specified. What is the situation of which the operator is required to be aware? In most of the experimental or field studies, it is not well defined or specific. Rather, and for obvious reasons, it is almost as though the investigators want the operator to be aware of 'anything in the environment which might be of importance if it should change unexpectedly'. This is, of course, not the phrase used in research. However, clearly, it is at the back of investigators' minds. The pilot of an aircraft needs a keen situation awareness because if anything abnormal occurs he or she must notice it and respond appropriately. However, there are, logically, an infinite number of events which may occur. Perhaps one, therefore, expects the operator to be aware of the status of just those variables which, if ever they change, represent significant events. But, what are those? If one does not define exactly what the set of events is that the operator must monitor, how can he or she devise an

> optimal or eutactic monitoring strategy? It is logically quite unreasonable to ask merely for 'situation awareness' as such, since if the set of events of which operators are to be aware is not defined, it is unreasonable to expect them to monitor the members of an undefined set; whilst if one defines a set, there is always the possibility that the dangerous event is not a member of the set, in which case the operators should not have been monitoring it. (p. 360)

This amorphous quality implies that SA may prove an inappropriate concept for considering moral blame or legal culpability. Moray and Inagaki's concerns, however, hint at a deeper problem again—and one that takes our focus here. SA in its canonical formulations and usual employment amounts to having a correct set of beliefs (about the relevant facts). As such, SA constitutes a status, an outcome resulting from the confluence of a wide array of factors, some originating inside and others outside the agent. SA does not connote an action, a practice, a role, a task, a virtue, or a disposition—all things that are the familiar object of moral appraisals.

2.2 SA and ethical theory

To examine this more closely, we turn to contemporary ethical theory and what it says about moral responsibility for holding a status, as distinct from performing an act or pursuing a goal. Following an orthodox categorization, we can divide ethical theories into three types: virtue theory; deontology; and consequentialism. Hearkening back to Aristotle, virtue theory focuses on people's character traits, asking whether they act courageously, wisely and benevolently. Virtue theories go on to provide guidance to people showing what acts they can perform to nurture the appropriate emotional dispositions, and so become virtuous. Deontology takes a different approach, focusing on specific types of actions—such as murder, theft, and lying (e.g. the "ten commandments"). Deontology prohibits these types of acts, and prohibits them irrespective of what effects they have on the person's character or on the world. The act itself, and not the virtue driving it, or the results of it, becomes the object of moral appraisal. As we can see, neither of these theories holds a person responsible for a status or an outcome. But what of consequentialism, the ethical theory that judges acts in terms of the amount of good they create (such as utilitarianism, which enjoins moral agents to create the maximum sum of happiness in the world)? Despite its name, consequentialism does not hold agents responsible for the outcomes of their actions.

Consequentialist ethics demands agents maximise *expected* overall utility; that is, they are enjoined to perform the act that, from their local perspective, probabilistically offers the greatest overall benefit in utility to the world (Breakey, 2009).

But why do these ethical theories, even those using consequences as the central organizing concept, resist assigning moral culpability to outcomes, statuses and achievements, and instead apply culpability to actions, decisions and virtues? We can find the answer in the well-known principle "ought implies can", which states, broadly speaking, that to be bound by a duty requires the possibility that one can knowingly perform that duty. Many theorists trace the principle to Immanuel Kant, but there is little need to invoke philosophical authorities when anyone can grasp the reasoning behind the idea. "Ought implies can" makes straightforward sense from both a retrospective and a prospective perspective. Retrospectively, it is morally confused (and alarming) to hold someone culpable for something that they could not have foreseeably prevented. If they could not have foreseeably prevented it, then they were not by their choice responsible for it; if they were not responsible for it, then they cannot be to blame for it. To blame someone for something they could not have done amounts to scapegoating them. Prospectively, it makes no sense to put forward a standard of conduct to which people can, despite all the best intentions and indeed despite performing the morally correct actions (given their local rationality, that is, their goals, knowledge, and constraints at the time of their act or decision) at every point, nevertheless fail to achieve.

2.3 The legal status of (loss of) SA

In order for criminal liability to apply, a person needs to perform negligent acts—that is, acts that a reasonable person, or ordinary professional actor, would not have performed. Recall the general definition of criminal negligence as doing something, or omitting to do something that is one's legal duty to do, showing wanton disregard for the lives and safety of others (Michaelides-Mateou & Mateou, 2010). Note the law here refers to actions and intentions, not statuses. In law, outcomes are not subject for (deontological) blame. We hold people responsible for performing or failing to perform acts because one can choose whether or not to perform acts, but cannot will into existence a particular status or outcome (such as accurate SA).

While we should therefore (as a matter of general principle) resist holding people responsible for having a particular status, our resistance can only increase when the particular status in question amounts to holding a false belief. Holding a false belief, like having incorrect or poor SA, has the additional difficulty that one usually does not (and sometimes could not possibly) know that one is in error. In her seminal paper on the topic, Mica Endsley was clear that SA comprised a “state of knowledge” (1995b, p. 36)—as opposed to the *process* of “situation assessment” (p. 36)—and well aware that one could lose SA without even knowing one had lost it: “A real issue concerns how people know when their SA is in error. Very often they may be completely unaware of how much they do not know or of the inaccuracy of their internal representation of the situation” (p. 57).

The ease with which terms like “error” and “mistake” can be applied to SA exacerbates the potential for wrongfully attributing blame. Consider three different ways we can use terms like ‘error’ and ‘mistake’ in cases relating to the possession of a false belief (all three usages can be observed in most publications on SA, see for example, Endsley (1995b)). First, a person might form their beliefs sloppily, or never search for evidence supporting or disproving those beliefs. In these cases, we might say they committed an error of evidence-gathering, logic, rationality, or reasoning. For example, someone might form their beliefs on the basis of the well-known gambler’s fallacy: ‘I concluded that the coin would fall heads this time on the basis of it falling tails the preceding five tosses.’ Second, when someone holds a false belief, we can express that as a “mistaken” or “erroneous” belief. “I thought my football team was going to win the premiership, but it turns out I was mistaken.” Third, when someone acts or makes a decision on the basis of a false belief, we might say that the act or decision was mistaken or in error. “I mistakenly locked my car door, because I believed I’d already taken the keys from the ignition and put them in my bag.” We can thus use the language of “mistake’ and “error” to apply to: actions-of-belief-formation; beliefs themselves; and actions-taken-on-the-basis-of-false-beliefs.

A person can make errors of reasoning and still come (through good fortune) to a correct belief. The coin-tosser in the foregoing example falls victim to the gambler’s fallacy in thinking the coin-toss is bound to come down heads, but she actually still has a decent chance (an even chance) of holding a true belief about the outcome of the toss. The converse holds true as well; a person can form beliefs in a wholly rational and epistemologically responsible manner, and still come to a false belief. Examples abound: if a cockpit malfunction gives a pilot an inaccurate reading on his or her altimeter, the pilot might make an entirely justified estimation of the airplane's altitude that departs from its real altitude. In philosophical terms, the pilot held a rational belief, notwithstanding that that belief turned out to be incorrect.

As a result, we need to be very careful in employing terms such as 'error' or 'mistake' in application to beliefs, or to plans made on the basis of beliefs. One can come to false beliefs without actually committing any sort of error of rationality or belief formation. One can also make decisions to act on the basis of false beliefs without committing any sort of error of rationality or means-end calculation. As a result, and as a trick of language, one can hold a 'mistaken' belief and engage in a "mistaken" plan of action without actually ever making a mistake in the ordinary sense—that is, without ever doing anything that one should (rationally) not have done. The sense of mistake that involves doing something one should not have done applies only to the first sense of mistake: a breach of rational or evidence-gathering norms during the process of belief acquisition. From our prior discussion of the principle of "ought implies can," one can only be culpable for this type of mistake.

Applied to SA, one can be culpable for mistakes in reasoning, monitoring or evidence-gathering: errors of belief-formation. However, when one fails to hold SA (or has poor SA) or one crafts mistaken plans on the basis of one's perceived awareness of the situation—'mistaken' not because of any rational or strategic error, but simply because the original perception failed to represent the true state of the world—one cannot be held to blame. One can only be culpable for the first type of mistake, in failing to perform the acts that are known to reliably lead to accurate belief formation; acts that the "ordinary skilled professional" would have carried out in order to update his or her SA. These are the actions that the Bolam test dictates. Possessing a particular state of mind—accurate SA—is a normative requirement that does not fit this bill.

2.4 Losing SA?

With this in mind, we need to be particularly careful when we speak of *failing* to hold SA, *losing* SA or failing to *maintain* SA. The use of these active verbs can imply a person is performing an action, and is therefore doing something for which they can be held morally or legally responsible. To the contrary however, a person can lose SA without acting at all, and even lose SA while assiduously carrying on appropriate monitoring tasks. Note, though, that in some cases a person knows they have lost SA and elects to do nothing about it, and continues to plan as if they had it. Here the loss certainly might be important to allocating blame, because we are speaking of something that they know has gone wrong, and of their failure to take appropriate actions given that fact.

Of course, one may to be to blame for the reasons why one does not possess SA. But here the usefulness of SA recedes, replaced by a concern with the epistemic rationality of specific actions. By "epistemic rationality" we mean the properties of actions that lead the actor towards particular outcomes. An agent might to be to blame because he or she failed to arrive at work sober, fell asleep, failed to look at a particular screen, did not double check a particular piece of information, or neglected to collaborate with a colleague. As specific actions that an agent knows they have to perform, all of these acts are in principle proper subjects of blame. However, there is no need in making judgments about these failures to invoke suboptimal outcomes (statuses) such as poor, inaccurate, or lost SA. This is particularly the case, as in looking back at an adverse event it is always possible to argue that someone's actions were not optimal; someone's awareness of the situation was not "complete" or "accurate". In hindsight, it is always possible to show that there was (what we now know to be a critical piece of) data available that the person in question did not observe. However, just as the ethical consequentialist does not hold people to blame for failing to achieve the very best outcome, but does hold them responsible if they fail to act in ways they knew would have probabilistically led to good outcomes, so too can an operator never be culpable for failing to maintain SA, but only for failing to perform "reasonable" actions—that is, the locally rational actions for a skilled professional in that situation.

We need to be wary, too, of using a lack of SA to cast about for unrealistic expectations about safety procedures. There may be good reasons in any particular case as to why people did, or neglected to do, what they did. With the benefit of hindsight, stating that a particular action or piece of information—which we now know to be crucial for a good outcome—was available to someone does not necessarily make him or her blameworthy for not acting or observing it in the moment. In order to deal with goal-conflicts and resource constraints, working professionals routinely work around or loosely interpret many official safety procedures to perform their jobs properly and effectively—a phenomenon discussed widely in human factors as the gap between work-as-imagined and work-as-performed. This underlies why 'work to rule' is an effective form of an industrial strike: namely, taking every single safety procedure in the rule-book in its full and literal meaning often suffices to gridlock production. Using SA in hindsight tempts prosecutors to search for reasons why SA was lost, and it will often turn out that professionals avoided one or other regulation. Here hindsight bias, the lure of SA as a normative standard, and overly officious and pedantic safety-guidelines conspire to form a perfect storm of mistaken culpability. In such cases, the test must always be what ordinary professionals actually do in such situations.

2.5 Conclusion

People can be held morally and legally responsible for their actions, their intentions, and their character. They are, however, not to be blamed for a certain status, especially a status like a false belief, which depends on an array of factors quite outside the agent's sphere of control, and which inescapably implies the agent does not know they possess that status (people holding a false belief do not know it is false).

Nothing we have said here denies that SA may be a useful and legitimate concept in designing systems and thinking about preventing accidents in medical and other domains. We might well be able to improve system performance and safety by asking: "In this situation, what information must a person possess in order to perform optimally?" So too, it may even make sense in the wake of failure to retrospectively ask: "In this given case, is there any actor who, if only they had known what was unfolding (i.e. if only they had possessed accurate SA), would have been able to prevent or mitigate this disaster?" Answering this question can be done without any claims that the lack of SA caused the accident in question, and (all the more) without any claims about how this particular person was not situationally aware in this case—for which there may be any number of valid reasons. While SA can therefore be a helpful concept in accident enquiries and investigations, even these contexts warrant some sensitivity in appealing to SA. Factual claims about "poor SA" or "mistaken SA" could be seized upon by prosecutors and the concept hijacked into a moral and legal discourse for which it is entirely unfit.

Ultimately, invoking SA becomes problematic when its discursive use broadens to include professional norms for behaviour, expecting a particular state of awareness from operators. When this normative requirement—to "maintain SA at all times"—is used to judge performance in questions of negligence and culpability, the object has exceeded both its appropriateness and usefulness. As such, we should purge SA from all moral and legal enquiries into blameworthiness.

PART II

-

THEORETICAL AND METHODOLOGICAL FRAMEWORK FOR ANALYSIS

Chapter 3:
The Foucauldian notion of discourse

Abstract

The concept of discourse is used across a number of disciplines, in different manners, and with different goals. For Foucault, discourses are constitutive; rather than reflecting an external world, our discursive practices produce our knowledges and understanding of reality. Discourses are sets of statements that create the objects of our knowledges and associate subject positions. As such, discourse provides the constraints for what can be meaningfully thought and said at a particular time and place. From this perspective, scientific social inquiry no longer needs to be concerned with a positivist search for the truth. Rather, with Foucault's notion of discourse, we can try understand how our knowledges, and the objects in them, are the contingent effects of our (scientific) discursive practices and how they come to be seen as meaningful or 'in the true'.

3.1 The broad scope of discourse

Over the last thirty years, 'discourse' has become a popular term in the social sciences and humanities. Originating in disciplines such as linguistics and semiotics, discourse is now frequently used in a variety of ways by different theorist from a wide range of academic disciplines: psychology, sociology, history, philosophy, anthropology, pedagogy, geography, political science, and many other (sub)disciplines (Howarth, 2000; Mills, 1997, 2003; Parker, 1992; Waitt, 2005; Willig, 2008). One of the reasons for the popularity of the concept of discourse, as well as its related notion of 'discourse analysis', can probably be found in social sciences' increasing dissatisfaction with positivist approaches to the study of complex social problems and cultures (Howarth, 2000; Waitt, 2005).

The concept of discourse is used in different disciplines, in different ways, with different meanings, and with different ways of analysing it. The broad range of application makes it difficult to provide a generally accepted definition of the notion. To circumvent this problem, as well as a sign of its frequent use, many researchers leave it undefined, assuming that the term discourse is an established commonsensical notion (Mills, 1997). The few researchers that do attribute their usage of discourse often refer to Michel Foucault. These attributions, however, are typically lacking in specificity—page numbers, (meticulously copied) quotes, or even referral to specific work or books of Foucault are omitted. In these cases, the attribution of discourse to Foucault is regarded as commonsensical and unproblematic as the usage of the notion of discourse itself (Sawyer, 2002).

To show the vast range of meaning, use, and scale of the notion of discourse, Vass (1992, p. 9, as cited by Titscher, Meyer, Wodak, & Vetter, 2000, p. 25) conducted a meta-study of the concept and provides the following list. Discourse can be any of the following:

1. (general): speech, conversation, discussion;
2. discursive presentation of a train of thought by means of a series of statements;
3. series of statements or utterances, chain of statements;
4. form of a chain of statements/expressions; the manner in which they came about (archaeology): scientific, poetic, religious discourse;
5. rule-governed behaviour that leads to a chain or similarly interrelated system of statements (=forms of knowledge) (medicine, psychology, etc.) (for instance in the work of Michel Foucault);
6. language as something practiced; spoken language (e.g. in the work of Paul Ricoeur);
7. language as a totality, the linguistic universe;
8. discussion and questioning of validity criteria with the aim of producing consensus discourse participants (e.g. in the work of Jorgen Habermas).

This list shows that the use and meaning of discourse ranges from something specific as a single utterance all the way to addressing the entire discursive universe. The latter, in turn, is regarded by some scholars as being synonymous with the entire (social) world (Howarth, 2000). Think, for example, of Derrida's famous quote: "*There is nothing outside of the text* [there is no outside-text; il n'y a pas de hors-text]" (Derrida, 1998, p. 158; italics in original). Or, as he said elsewhere, "when language invaded the universal problematic … everything became discourse" (Derrida, 1978, p. 280). Discourse theorists were able to make such bold statements about the importance of discourse, as they assumed that all social phenomena have meaning, and that such phenomena could thus be seen as 'symbolic systems of practice' (Howarth, 2000). Discourse, then, acquires a central role; it constitutes these symbolic systems of meaning.

In understanding the expansion of scope from a narrow linguistic approach to a larger more sociological approach, the work of Foucault is particularly important—which is probably the reason that he is often, but incorrectly (see Sawyer, 2002), named as the father of the concept of discourse. Briefly, in his earlier, archaeological work, Foucault sees discourses as historically contingent and rule-governed systems that govern what can meaningfully be said and thought at a particular time and place. In his genealogical work, Foucault adjusts the angle of his studies, and starts to look into the question of how discourses change, thereby turning to notions as power, subjectivity, and the role of institutions. "Rather than describing the historical rules that make possible sets of statements, Foucault is now concerned with the way in which discourses are shaped by social practices and the way they in turn shape social relationships and institutions" (Howarth, 2000, p. 8). The point here is that for Foucault discourse is no longer purely a linguistic concept, rather he understands discourse as discursive practices that produce our (social) world.

3.2 The fluctuating meaning of Foucault's notion of discourse

So far, we have seen that Foucault enlarges the scope of discourse, however, we still do not know what Foucault means and wants with his concept of discourse. Unfortunately, Foucault never provides a single theory focussing on the term discourse, but develops the concept in wide range of books, articles, lectures, and interviews. This makes it hard to 'pin down' a single understanding of discourse, and can probably also explain part of the difficulty scholars have in defining the concept. In his book, 'The Archaeology of Knowledge' (Foucault, 2002a), after ninety pages of trying to explain the analysis of a number of 'discursive regularities'—such as discourse and 'discursive formations' (see chapter 4)—Foucault realises that he has not been successful in explaining what he means by discourse:

> … instead of gradually reducing the rather fluctuating meaning of the word 'discourse', I believe that I have in fact added to its meanings: treating it

> sometimes as the general domain of statements, sometimes as an individualizable group of statements, and sometimes as a regulated practice that accounts for a certain number of statements … (Foucault, 2002a, p. 90)

If, with Foucault (2002a)[7], we accept that 'statements' simply refer to spoken utterances or written sentences that have meaning, then Foucault thus gives us three directions in which he uses the concept of discourse in his work. First, he sees discourse as 'the general domain of statements', the general domain of all written or spoken language that has meaning. Foucault uses this broad definition particularly in his earlier work when he talks about 'discourse' at a larger theoretical level. Second, when Foucault refers to discourse as 'an individualizable group of statements', he seems to talk about a more coherent and identifiable group of statements, a more or less discrete unity of discourse, which often refers to as 'discursive formations' (Foucault, 2002a). Examples of this are the 'discourse of psychopathology in the nineteenth-century' or the 'medical discourse' (see next chapter). Thirdly, when Foucault is describing discourse as 'a regulated practice that accounts for a certain number of statements', he seems to be referring to the regularities and rules that make up a particular discursive formation.

The difficulty that comes with Foucault's concept of discourse is that he uses these three meanings interchangeably, without specifying which particular meaning he is using. For this reason Howarth (2000), in his book '*Discourse*', calls Foucault's conception of discourse "paradoxical" (p.48). He explains that even though the concept plays such a pivotal role in Foucault's entire oeuvre, "the concept remains frustratingly unclear" (p. 48). Howarth depicts Foucault's conception of discourse by sketching two 'extremes' of what the term refers to, which add to the three interpretations that I have discussed above:

> At one extreme, discourses are autonomous systems of rules that constitute objects, concepts, subjects and strategies, thereby governing the production of scientific statements.
>
> …

[7] Later in 'The Archaeology of Knowledge' Foucault problematizes this conception of statement. He then describes the statement not as a linguistic or logical unit, not even as 'speech act', but as an 'enunciative function' that enables groups of signs to exist (Foucault, 2002a). However, for this thesis' purposes the conception of statements as text that has meaning holds.

> At the other extreme, as against this constitutive view of discourse, Foucault (1979a: 101–2) argues that discourses are 'tactical elements or blocks operating in the field of force relations'. From this 'strategical' perspective, discourses are the means for different forces to advance their interests and projects, while also providing points of resistance for counter-strategies to develop. (Howarth, 2000, pp. 48-49)

Howarth (2000) subsequently connects these different conceptions of discourse with different periods of Foucault's writing. He sees the first 'extreme' as the narrower, more formative conception of discourse as relating to Foucault's archaeological projects. The other extreme represents a conception of discourse, which Foucault deploys in his genealogical investigations of power, knowledge, and subjectivity. Especially these associations of discourse with relations of power and society might explain the wide range of interest and resulting popularity of Foucault's conception of discourse. Let us now take a look how Foucault came to such a conception.

3.3 Discourse versus ideology

The political events of May 1968 in Paris, when the student unrest provoked a broader societal disturbance, illustrated the need for new approaches to political change. Many academics, including Foucault, were drawn into the campaigns and openly joined the gay liberation movement, the prison movement, the ecological movement, as well as the demonstrations for workers' and women's rights (Arribas-Ayllon & Walkerdine, 2008; Mills, 1997; O'Farrell, 2006)[8]. The academics that took part in these movements questioned the liberal humanist and conservative orthodoxies, as well Marxist theories. According to them, the new order of social problems seemed to lie beyond these out-dated perspectives. Foucault's conceptualisation of discourse seems to be founded in this frustration and provides a new direction for moving beyond these models, dominated by Marxist and humanist epistemologies (Arribas-Ayllon & Walkerdine, 2008).

---

[8] Interestingly, O'Farrell (2006, p. 190) notes: "France did not develop any active left wing terrorist groups – a feature of the landscape elsewhere especially Germany and Italy in the 1970s. There were some suggestions at the time that it was the moderating influence of Foucault and other intellectuals which helped contribute to this state of affairs in France."

In a nutshell, a simple Marxist model sees 'ideology' as certain statements and ideas, legitimised by (governmental) institutions, which produce in individuals a (conceptual) reality that is not in their best interests. Marxist scholars thus see ideology as something negative and oppressive, as a set of false (misleading) beliefs—a 'false consciousness'. Subsequently, these false beliefs are being instilled and exploited by society's more powerful (institutions), for example to control the working classes. Marxism thus offers a clear political agenda, as it calls for a revolution, a remodelling of the economy, an alleviation of the oppression of the working classes, by revealing the illusionary effects of ideology (Mills, 1997).

Foucault's notion of discourse lacks this political agenda, but aims to explain how people are complicit in their own oppression, without assuming that people are simply passive victims of systems of thought—of ideology (Howarth, 2000; Mills, 1997). For Foucault there is no conspiracy, no one (social group, institution, or government) pulling the strings and trying to trick people into doing things that are not in their best interests. Rather, for Foucault there are discourses, (power) relations, 'technologies', and other 'apparatuses' and mechanisms, constituting and maintaining certain practices, in which everybody takes part.

At a more conceptual level, Foucault explains some of 'difficulties' he has with the concept of ideology.

> The notion of ideology appears to me to be difficult to make use of, for three reasons. The first is that, like it or not, it always stands in virtual opposition to something else which is supposed to count as truth. Now I believe that the problem does not consist in drawing the line between that in a discourse which falls under the category of scientificity or truth, and that which comes under some other category, but in seeing historically how effects of truth are produced within discourses which in themselves are neither true nor false. The second drawback is that the concept of ideology refers, I think necessarily, to something of the order of a subject. Thirdly, ideology stands in a secondary position relative to something which functions as its infrastructure, as its material, economic determinant, etc. For these three reasons, I think that this is a notion that cannot be used without circumspection (Foucault, 1980, p. 118)

Based on this quote, I will develop Foucault's conception of discourse by looking at how it differs from that of ideology.

3.4 Discourse and truth

3.4.1 Discourse and reality

Foucault's first problem with ideology is the idea that it always stands in opposition to some kind of truth (Foucault, 1980). As ideology is characterised as a false consciousness or imagined representation of reality, it assumes that there is a consciousness that is not false or not imagined—but real. The position from which to critique this false consciousness must therefore necessarily stand outside ideology in order to make this a-perspectival (objective) claim about the 'truth' of reality. Foucault, however, does not position himself outside the systems of thoughts, knowledges, ideas, and practices that he is analysing.

Where Marxist theorist see language as a medium to make people believe in a false reality that is not in their own best interest, Foucault does not assume that there is a straightforward relation between language and reality, as discourses do not simply translate reality into language. Rather than seeing discourses as reflecting or mirroring reality, he sees discourse as constitutive of reality. Foucault, however, is not saying that 'there is nothing outside discourse'—as some of his renowned contemporaries did (see above)—the extreme relativist position that everything is (socially) constructed and that there is no pre-discursive. When Foucault says that "nothing has meaning outside of discourse" (Foucault, 1972, as read in Waitt, 2005, p. 170) he does not deny that material objects and social practices exist. Foucault does not deny that objects exist externally to thought, rather he says that they cannot constitute themselves as objects outside discourse. In other words, reality is categorised and understood through the discursive structures available to us. We can only think and experience the physical world through discourse, its categories, and the structures it imposes on our thinking.

This epistemological position, which suggests that things can have a real, material, existence in the world, but that they only take on meaning and become objects of knowledge through discourse has been labelled the constructionism (Hall, 2001; Willig, 2008). However, Foucault is not concerned with the ontological or epistemological status of discourse. As we have seen, Foucault is not interested in ‘truth’, but rather “in seeing historically how effects of truth are produced within discourses which in themselves are neither true nor false” (Foucault, 1980, p. 118).

> One of the misconceptions concerning Foucault’s work is that he ignores concrete reality at the expense of so called discourse. According to this misconception, things only acquire reality as the result of social practices or the way we talk about them. Foucault, contrary to this, holds that there is in fact an intractable physical reality – but the way we describe, interact with and focus on it is highly variable and by no means fixed. The only way we can apprehend this raw level is by means of a whole panoply of complex cultural and conceptual tools which differ considerably according to historical period and culture. Foucault argues that the way we link words and things is by no means obvious, and that there is simply no way of pronouncing any of the links we make between words (or knowledge) and things to be absolutely true for once and for all. (O'Farrell, 2006, p. 188)

As O’Farrell explains, for Foucault the world is not organised or categorised in one necessary way: discourses do not reflect reality, but reality is constituted through (historically contingent) categorisations and descriptions. In seeing ‘discourse as practice,’ Foucault (2002a, p. 52) argues that “[w]hat, in short, we wish to do is to dispense with ‘things’”:

> To substitute for the enigmatic treasure of ‘things’ anterior to discourse, the regular formation of objects that emerge only in discourse. To define these *objects* without reference to the *ground*, the *foundation of things*, but by relating them to the body of rules that enable them to form as objects of a discourse and thus constitute the conditions of their historical appearance. (Foucault, 2002a, pp. 52-53, italics in original)

As such, Foucault sees discourses “no longer … as groups of signs (signifying elements referring to contents or representations) but as practices that systematically form the objects of which they speak” (Foucault, 2002a, p. 54). Instead of ‘representing’ some external reality, discursive practices systematically form the object of which they speak: medical and psychiatric discourses produce the mentally ill (and specific psychopathologies); penological discourses produce the criminal; discourses on sex produce sexuality. Foucault, for example, explains that the "psychiatric discourse finds a way of limiting its domain, of defining what it is talking about, of giving it the status of an object – and therefore of making it manifest, nameable, and describable" (Foucault, 2002a, p. 46). This, however, doesn't mean that before these discourses there was no mental illness, crime, and sexuality. There was no 'void', we just did not understand and experience the world in these categories.

3.4.2 Discourse and speaking 'in the true'

In 'The Order of Discourse', his inaugural lecture at Collège de France (December 1970), Foucault explains:

> [W]e must not resolve discourse into a play of pre-existing significations; we must not imagine that the world turns towards us a legible face which we would have only to decipher; the world is not the accomplice of our knowledge; there is no prediscursive providence which disposes the world in our favour. We must conceive discourse as a violence we do to things, or in any case a practice which we impose on them; and it is in this practice that the events of discourse find the principle of regularity (Foucault, 1981, p. 67)

Foucault thus argues that the regularities we see in reality are brought about by the regularities of discourse, which we 'violently' impose on reality. Discourse, thus, should be seen as something that constrains our perceptions, our thoughts, our reality—our 'truth'. "Truth is a thing of this world: it is produced only by virtue of multiple forms of constraint" (Foucault, 1980, p. 131).

Foucault extensively speaks about the various types of constraints of discursive practices in 'The Order of the Discourse' (1981). He explains that there are a number of 'procedures' to constrain discourses in order to 'ward off its power and dangers', to control what is seen as meaningful (and consequently what is seen as nonsensical), what counts as knowledge, and what people are able to think at a certain time (epoch) and place.

> [I]n every society the production of discourse is at once controlled, selected, organized and redistributed by a certain number of procedures whose role is to ward off its power and dangers, to gain mastery over its chance events, to evade its ponderous, formidable materiality (Foucault, 1981, p. 52).

These procedures, or constraints, can be grouped into three groups. First, Foucault (1981) explains, there are external procedures or 'systems of exclusion' that operate on discourses and determine what can meaningfully be said when, where, and by whom. Secondly, there are internal constraints, internal 'procedures of rarefaction' as Foucault calls them, as discourses have an internal structure that enables certain constructions and renders others meaningless (and powerless). Finally, Foucault believes there is a third group of procedures that permit the control of discourses, which is neither external nor internal to discourse, but rather is about "determining the condition of their application, of imposing a certain number of rules on the individuals who hold them, and thus of not permitting everyone to have access to them" (Foucault, 1981, p. 61).

> All these external and internal procedures for constraining—controlling, selecting, organizing, and redistributing—discourses work towards one thing: the production of discourses. As these procedures all work towards the delimitation of what counts as knowledge, what can be said and thought at a particular time, for Foucault, this explains the homogeneity of discourses from a particular time. It is not because everybody simply agreed on a particular representation of something, but because all of the (sanctioned) discourses are produced with similar constraints (Mills, 1997).

One of such 'systems of exclusion' is our 'will to truth'. According to Foucault, around the time of Plato a certain division was introduced that separated true discourse from false discourse. This division gave rise to our 'will to truth', the will to produce true representations, to speak or write the truth (Foucault, 1981). Since this Platonic division, we assume that the separation of true and false, as well as our concern of with representing the truth, is necessary and self-evident. In our writing and speaking we are concerned with presenting accurate (true) 'copies' (using Plato's vocabulary) of our external reality. This 'will to truth', the search for 'vraisemble', is also the main goal for academia: "Western literature sought to ground itself on the natural, the 'vraisemble', on sincerity, on science as well – in short, on 'true' discourse" (Foucault, 1981, p. 55).

> For Foucault (1981), this will to truth, as are other 'systems of exclusion', is supported and reinforced by institutions. "This will to truth, like the other systems of exclusion, rests on an institutional support: it is both reinforced and renewed by whole strata of [institutional] practices" (p. 55), such as universities, schools, publishers, libraries, scientific institutions, government bodies, etc. "But it is also renewed, no doubt more profoundly, by the way in which knowledge is put to work, valorised, distributed, and in a sense attributed, in a society" (p. 55). All the institutions, and institutionalised practices, work to promote and reinforce discourses that they classify as true, whereas they exclude discourses which they characterise as false. For Foucault, only discourses that are 'in the true' will be promoted: "It is always possible that one might speak the truth in the space of a wild exteriority, but one is 'in the true' only by obeying the rules of a discursive 'policing' which one has to reactivate in each of one's discourses" (Foucault, 1981, p. 61).

Foucault illustrates the importance of being 'in the true', as compared to 'speaking the truth', using a clear and convincing example:

> People have often wondered how the botanist of biologists of the nineteenth century managed not to see that what Mendel was saying was true. But it was because Mendel was speaking of objects, applying methods, and placing himself on a theoretical horizon which were alien to the biology of his time. … This was a new object which called for new conceptual instruments and new theoretical foundations. Mendel spoke the truth, but he was not 'within the true' of the biological discourse of his time: it was not according to such rules that biological objects and concepts were formed. It needed a complete change of scale, the deployment of a whole new range of objects in biology for Mendel to enter into the true and for his propositions to appear (in large measure) correct. (Foucault, 1981, pp. 60-61).

So, where Mendel was speaking the 'truth' (in line with our current believes), he was not 'in the true'. Foucault (1981, p. 61) contrasts Mendel's case by saying that about thirty years earlier, another botanist, Scheiden, “denied plant sexuality,” however, as he did so “in accordance with the rules of biological discourse, [he] was merely formulating a disciplined error”. These historically contingent mechanisms that govern the processes of inclusion and exclusion, that delimit what can truthfully be said and done, are determined by what Foucault termed a ‘regime of truth’[9].

> Each society has its regime of truth, its ‘general politics’ of truth: that is, the types of discourse which it accepts and makes function as true; the mechanisms and instances which enable one to distinguish true and false statements, the means by which each is sanctioned; the techniques and procedures accorded value in the acquisition of truth; the status of those who are charged with saying what counts as true. (Foucault, 1980, p. 131)

Vital to a regime of truth is the mutually reinforcing relationship between power and knowledge. “Truth isn’t outside power, or lacking in power” (Foucault, 1980, p. 131). Power, in a Foucauldian conception, should not be understood in the conventional manner as something that a particular individual or group of agents has, and who is consequently able to enforce it upon those who do not have power. Instead of conceptualising power in negative terms, as a repressive, hierarchical, local source, for Foucault power is a positive, productive force that is everywhere.

> If power were never anything but repressive, if it never did anything but to say no, do you really think one would be brought to obey it? What makes power hold good, what makes it accepted, is simple the fact that it doesn’t only weigh on us as a force that says no, but that it traverses and produces things, it induces pleasure, forms knowledge, produces discourse. It needs to be considered as a productive network which runs through the whole social body, much more than as a negative instance whose function is repression. (Foucault, 1980, p. 119)

[9] Here I have explicitly entered Foucault’s genealogical—rather then archaeological—approach. However, instead of rigidly avoiding it, in order to work from a ‘pure’ archaeological perspective, I see the two approaches as complimentary and mutually reinforcing. Because Foucault’s conception of discourse crosses both approaches, this chapter occasionally touches upon genealogical notions of truth, power, and the subject. Moreover, as chapter 8 explicitly presents an archae-genealogical approach to study the object of safety culture, it will draw on a number of these central genealogical concepts for analysis.

Power is thus productive. It produces discourse, knowledge, and truth, which in turn produce or sustain certain power relations: "power and knowledge directly imply one another; that there is no power relation without the correlative constitution of a field of knowledge, nor any knowledge that does not presuppose and constitute at the same time power relations" (Foucault, 1991a, p. 27). For this mutually constitutive function Foucault coined the analytical term 'power-knowledge relation', which he uses to stress that knowledge and discourse do not operate in a social void (Foucault, 1980).

For this section's purposes, it is important to understand that for Foucault questions about the absolute truth of knowledge (or discourses and the objects in them) are not important. It is the processes by which a particular discourse attains (and functions with the status of) truth that interests him. "I want to try discover how this choice of truth, inside which we are caught but which we ceaselessly renew, was made – but also how it was repeated, renewed, and displaced" (Foucault, 1981, p. 70).

In stark contrast with earlier Marxist theorists, who critique ideology as a false consciousness from a position of truth, Foucault is not claiming to speak from a position of truth. Foucault is also not saying that nothing is true and that what is designated as truth simply serves the interests of power. Foucault understands that he can only understand and critique certain systems of thought from within the boundaries of the discursive structures that make up reality. There is thus no a-perspectival position to speak from, to critique certain discourses from, or to make truth claims from, there is just discourse. As Foucault eloquently explains:

> It [archaeology] is not critical, most of the time; it is not a way of saying that everyone else is wrong. It is an attempt to define a particular site by the exteriority of its vicinity; rather than trying to reduce other to silence, by claiming that what they speak is worthless, I have tried to define this blank space from which I speak, and which is slowly taking shape in a discourse that I still feel to be so precarious and so unsure. (Foucault, 2002a, pp. 18-19)

3.5 Discourse and the subject

As a second problem with the Marxist concept of ideology, for Foucault, was that it refers, to "something of the order of a subject" (Foucault, 1980, p. 118).

> If there is one approach that I do reject, however, it is that (one might call it, broadly speaking, the phenomenological approach) which gives absolute priority to the observing subject, which attributes a constituent role to an act, which places its own point of view at the origin of all historicity – which, in short, leads to a transcendental consciousness. (Foucault, 2002b, p. xv)

In his work Foucault explicitly tries to avoid relying on the phenomenological subject, because it assumes a founding human subject that serves as the origin of thought and discourse (Howarth, 2000). Mills explains about Foucault's refusal to refer to the sovereign subject as a unitary being: "Foucault is very much part of post-structuralist thinking, which questioned the very fundamental bases of liberal humanist ideology, rooted as it is in the notion of the individual self with agency and control over itself" (2003, pp. 33-34). Instead of accepting the seemingly self-evident psychological and sociological categories as the subject and the self, particularly in his genealogical projects, Foucault is concerned with uncovering the power-knowledge relations invested in the use of such categories:

> But this historical contextualization needed to be something more than the simple relativisation of the phenomenological subject. I don't believe the problem can be solved by historicizing the subject as posited by the phenomenologists, fabricating a subject that evolves through the course of history. One has to dispense with the constituent subject, to get rid of the subject itself, that's to say, to arrive at an analysis which can account for the constitution of the subject within a historical framework. And this is what I would call genealogy, that is, a form of history which can account for the constitution of knowledges, discourses, domains of objects etc., without having to make reference to a subject which is either transcendental in relation to the field of events or runs in its empty sameness throughout the course of history (Foucault, 1980, p. 117).

Foucault is thus more interested in the processes that constitute the subject within a historical framework. As such, he chooses to investigate how discourses constitute knowledges that enable particular subjectivities. Just like the objects of knowledge, discourse also produces the subjects of knowledge. Discourse produces figures that personify the particular forms of knowledge as defined by the discourse (Hall, 2001). In other words, discourse makes available certain positions for the enunciator, reader, or viewer to take up in relation to that discourse (Henriques, Hollway, Urwinn, Venn, & Walkerdine, 2002). Such positioning "involves the construction and performance of a particular vantage point … ; it offers not only a perspective form which to view a version of reality, but also a moral location within spoken interaction" (Arribas-Ayllon & Walkerdine, 2008, p. 102).

The medical discourse, for example, constitutes the doctor as a particular subject. When positioning himself within the medical discourse by assessing a patient in need of medical help, the doctor also creates the subject position for himself as a person with a particular status, authority, knowledge, skills, moral and ethical obligations. Another position created by the medical discourse, is that of the patient, "which locates them as the passive recipient of expert care within a trajectory of cure" (Willig, 2008, p. 113). Other examples of subjects that Foucault was particularly interested in are the madman, the homosexual, and the individualised criminal.

Foucault does not deny that every discourse has a subject—in the sense of a discursive source—however, for Foucault this subject is not a discursive source, not a transcendental "speaking consciousness, not the author of the formulation, but a position that may be filled in certain conditions by various individuals" (Foucault, 2002a, p. 129). Just like objects of knowledge, for Foucault (2002a), discourses also produce subjects of knowledge:

> I showed earlier that it was neither by 'words' nor by 'things' that the regulation of the objects proper to a discursive formation should be defined; similarly, it must now be recognized that it is neither by recourse to a transcendental subject nor by recourse to a psychological subjectivity that the regulation of its enunciations should be defined. (Foucault, 2002a, pp. 60-61)

In short, subjects do not produce discourses, but rather, subjects are the effects of discourses.

> In this way Foucault challenged conventional Western notions of the subject as an autonomous and sovereign entity, an individual who is fully endowed with

> consciousness and independent source of meaning. Instead, following poststructuralist thinking, the subject cannot exist outside discourse. That is, people take up particular social identities through their own particular gendered, classed, sexualised, and ethnic subject-positions produced within discourse. (Waitt, 2005, p. 173)

In stepping away from the constituting (epistemic) subject, Foucault also offers an alternative to the epistemological 'problem of knowledge'. Briefly, the problem of knowledge—commonly referred to as the question 'how do we know what we know?'—arises when the human mind turns back on itself and reflects on its own operations (Kologlugil, 2010). Traditionally, the problem of knowledge is posed as a problem of the relation between on the one hand the knowing epistemic subject, based on the Cartesian cogito, and on the other hand objective reality, which exists out there independently of the knowing subject. The problem of knowledge for epistemology thus lies in the dualist (ontological) divide between the subject and object of knowledge, which exist independently of each other, making it hard to sustain the claim that we can acquire knowledge (Gergen, 1999; Kologlugil, 2010). However, "in constructing his cogito, Descartes was not only giving an answer to the epistemological problem; he was also defining the problem itself" (Kologlugil, 2010, p. 5). In other words, the epistemic problem of knowledge lies in the formulation of the question; "the insoluble problem of knowledge is only insoluble because of the dualist metaphor used to define the problem" (Gergen, 1999, p. 11). There is no necessity that calls for invoking a subjective human consciousness in addressing the problem of knowledge.

With his archaeological conception of discourse, Foucault frees us from the Cartesian problem of knowledge by drastically redefining it. He does not pose the problem in reference to a (epistemological) knowing subject and does not want to search for universal criteria of true knowledge, but rather looks for the discursive regularities and rules that make knowledge itself possible in a certain society (Kologlugil, 2010). The problem is thus no longer to describe how a subject acquires true (objective) knowledge, but to understand how a particular discourse—body of knowledge—acquires the status of truth.

Foucault's opposition to searching for origins of ideas in the minds of individuals also sets it apart from Marxist conceptions of ideology.

> I think I would distinguish myself from both the Marxist and the para-Marxist perspectives. As regards Marxism, I'm not one of those who try to elicit the effects of power at the level of ideology. … Because what troubles me with these analyses which prioritise ideology is that there is always presupposed a human subject on the lines of the model provided by classical philosophy, endowed with a consciousness which power is then though to seize on. (Foucault, 1980, p. 58)

3.6 Determinants of discourse

Finally, as a third problem with ideology, Foucault saw it as standing "in a secondary position relative to something which functions as its infrastructure, as its material, economic determinant, etc." (Foucault, 1980, p. 118). In stark contrast with Marxism, Foucault did not believe in an economic determinant of discourse, in that it would govern what could be said and thought at a particular place and time (Mills, 1997).

Foucault explains that the traditional 'history of ideas', as Foucault labels his own academic discipline, looks for causal relations between discourse and political events, economic phenomena, and institutions. "A causal analysis … would try to discover to what extent political changes, or economic processes, could determine the consciousness of scientists", and would thus try to discover how "at a period in which industrial capitalism was beginning to recalculate its manpower requirements, disease took on a [sic] social dimension" (Foucault, 2002a, p. 180). Foucault, however, is concerned with the 'conditions of possibility' for certain discourses, as he "wishes to show not how political practice has determined the meaning and form of medical discourse, but how and in what form it takes part in its conditions of emergence, insertion, and functioning (Foucault, 2002a, p. 181)".

In his archaeological work, Foucault studies discourses as systems of statements in order to examine the historically and culturally specific rules that determine what can meaningfully be thought and said in society at particular times and places. The discursive rules and structures that Foucault is interested in, however, are not directly determined by external sociological or economical factors, but rather originate in the complexity of discourse itself.

> … what we are concerned with here is not to neutralize discourse, to make it the sign of something else, and to pierce through its density in order to reach what remains silently anterior to it, but on the contrary to maintain it in its own consistency, to make it emerge in its own complexity. (Foucault, 2002a, p. 52)

In outlining his intentions for his archaeological approach to discourse, which will be more elaborately discussed in the next chapter, Foucault explains that he "does not treat discourse as *document*, as a sign of something else", but rather than he "is concerned with discourse in its own volume, as a *monument*. It is not an interpretative discipline: it does not seek another, better-hidden discourse. It refuses to be 'allegorical'" (Foucault, 2002a, p. 155). This approach to discourse thus distances itself from hermeneutical approaches that try to uncover, through interpretation, the underlying meanings of discourses and social practices. For Foucault, there is no hidden meaning, concealed from the subjects of discourse, at a deeper level than discourse itself. Foucault also warns against essentialism. He wants to move away from seeking underlying essences and unities in discourse and move towards an anti-reductionist approach in which discourse is described in its own terms (Howarth, 2000; Mills, 1997). "We sought the unity of discourse in the objects themselves, in their distribution, in the interplay of their differences, in their proximity or distance—in short, in what is given to the speaking subject" (Foucault, 2002a, p. 51).

Foucault is not chiefly concerned with the influence of external factors on the content of discourses. Rather, he want to analyse how a particular discourses "as a practice concerned with a particular field of objects, finding itself in the hands of a certain number of statutorily designated individuals, and having certain functions to exercise in society, is articulated on practices that are external to it" (Foucault, 2002a, p. 182). Instead of assuming, as ideology does, a simple relationship between discourse and economic or social forces, Foucault "saw the relationship between economics, social structures and discourses as being a complex interaction with none of the terms of the equation being dominant" (Mills, 1997, p. 36). Economic forces are thus just one type of relations influencing discourses. Discourses, in their turn, enable or sustain certain economical and political systems, as well as many other social structures. As an example, Foucault "wishes to show not how political practice has determined the meaning and form of medical discourse, but how and in what form it takes part in its conditions of emergence, insertion, and functioning (Foucault, 2002a, p. 181)".

3.7 Conclusion

In seeing 'discourse as practice,' Foucault (2002a) outlines his intentions for this currently popular notion. In explaining that "to speak is to do something" (2002a, p. 230), Foucault sees discourse as the discursive practices that produce our reality. Discourse does thus not reflect reality, but rather, reality and truth are constituted through the regularities of discourse. Foucault also rejects accounts of discourse that assume a founding human subject that serves as the origin of discourse, as an *a priori*. Discourses are thus sets of statements that create the objects of our knowledge and associate subject positions. Discourse provides the constraints for what can be meaningfully thought and said at a particular time and place.

Foucault's point, however, with the notion of discourse, is not an epistemological one; Foucault is not arguing that there is no such thing as truth, as he is not engaged in a relativist rejection of truth. Rather, he wants to understand how our knowledges, and the objects in them, are the contingent effects of our (scientific) discursive practices and how they come to be seen as meaningful or true. From this perspective, scientific social inquiry no longer needs to be concerned with a positivist search for the truth. By aiming to describe the historically contingent rules and regularities that are internal to discourse, Foucault tries to question the body of knowledge that people would accept as self-evidently true. In outlining Foucault's archaeological approach, the next chapter will also further refine his conception of discourse as well as the analytical use he has for it.

By outlining Foucault's intentions with his notion of discourse, this chapter has aimed to provide the background and elaborate on some of the assumptions for the various analytical studies presented later in this thesis. These individual studies will briefly review these ideas for their specific purposes.

Chapter 4:
Foucault's archaeological approach
& the formation of objects

Abstract

In his archaeological projects, Foucault aims to show that the constraints on our discourses are historically contingent and thus open for examination and possible change. He argues that these constraints are not external but rather that the regularities that characterise our discursive practices are internal to discourses themselves. By treating discourses as monuments, rather than documents, the archaeological level of description aims to describe the systematic regularities intrinsic to discourse. Archaeology can be thus seen as particular level of description, rather than an mechanistic method. In his various archaeological studies, Foucault provides a number of analytical concepts for archaeological analysis. He explains that the unity of discursive formation can be described through the 'rules of formation' for its various elements. In focusing on the formation of objects, Foucault specifies the 'surfaces of emergence', the 'authorities of delimitation', the 'grids of specification' of objects, as well as the 'discursive relations' among these aspects. Alternatively, Foucault suggests capturing the 'conditions of possibility' for thought at a particular time and place.

4.1 The aim of Foucault's archaeological projects

In his first archaeological work, originally 'Folie et déraison' from 1961 but later translated and published as 'Madness and Civilization', Foucault (2001b) is unwilling to buy into the commonly accepted idea that our modern understanding of madness—as mental illness, for which the modern disciplines of psychology and psychiatry offer the possibility of a scientific cure—is the result of (scientific) progress and presents a progressively humane manner to deal with madness. Foucault (2001b) does not consider the change from physical confinement of the mad in the 'Hospital General' during the Classical Age towards the 'more humane' modern treatment of the mad—the mentally ill—in the asyslum, as envisioned by reformers as Tuke and Pinel, as the liberation of the mad from the cruelties of the past:

> We must therefore re-evaluate the meanings assigned to Tuke's work: liberation of the insane, abolition of constraint, constitution of a human milieu—these are only justifications. The real operations were different. If fact Tuke created an asylum where he substituted for the free terror of madness the stifling anguish of responsibility; fear no longer reigned on the other side of the prison gates, it now raged under the seals of conscience. (Foucault, Michel, 2001b, p. 234)

Foucault (2001b) sees different ages as having their own 'experiences' of madness, that need to be understood in their own historical context—including our modern psychiatry. Foucault explains about the emergence of the 'the mad' as an object of knowledge (who were part of a larger group, 'the idle', who represent 'unreason'):

> … they chose a group that to our eyes is strangely mixed and confused. But what is for us merely an undifferentiated sensibility must have been, for those living in the classical age, a clearly articulated perception. It is this mode of perception which we must investigate in order to discover the form of sensibility to madness in an epoch we are accustomed to define by the privileges of Reason. (Foucault, Michel, 2001b, p. 42)

With his analysis, Foucault intends to question the modern idea that 'mental illness' is a natural, value-free, objective, or 'true' category—discursive object—that came out of these new, more humane, insights: "The asylum of the age of positivism, which it is Pinel's glory to have founded, is not a free realm of observation, diagnosis, and therapeutics; it is a judicial space where one is accused, judged, and condemned" (Foucault, Michel, 2001b, pp. 255-256). Rather than having humanistic reasons, Foucault sees the modern understanding of madness as mental illness, as well as its related scientific domain of modern psychiatry, as a discursive construction of the modern age, which was the contingent effect of some historical events and conditions of possibility. Foucault shows, for example, the relationship between the practice of enclosing the 'mad' and the establishment of a new social order (see Foucault, Michel, 2001b).

The emergence of the psychiatric discourse—and its associated practices and objects—was not the necessary or logical end-point of (inevitable) human progress, but merely one possible result of a complex network of relations between (historical) events. The point that Foucault occasionally makes is larger body of work is that much of our highly-valued progress of both science and our modern society is the accidental result of some unrelated changes or events (Kendall & Wickham, 2003). Foucault's conception of history is thus explicitly anti-teleological, as it does not involve any assumptions of progress—nor regress. His perspective breaks with a linear and totalizing conception of history, which searches for origins and takes the human subject as the basis for understanding history. Rather, with his archaeological projects, Foucault is interested in providing:

> ... a history which is not that of its growing perfection, but rather preface that of its *conditions of possibility*; in this account, what should appear are those configurations within the space of knowledge which have given rise to the diverse forms of empirical science. Such an enterprise is not so much a history, in the traditional meaning of that word, as an 'archaeology'. (Foucault, 2002b, pp. xxiii-xxiv; italics are mine)

With his various archaeological projects, Foucault aims to reveal the contingency of our knowledges, as well as our inability to think beyond the categories (and 'truths') of our current discourses. Foucault explains, "I want to try discover how this choice of truth, inside which we are caught but which we ceaselessly renew, was made – but also how it was repeated, renewed, and displaced" (1981, p. 70). The goal of his archaeological approach is thus not to make any judgements about the truth-value of discourses, but rather to investigate the discursive framework—the support structure that relates it to other statements and gives it meaning and value—that keep these discourses in place, that makes them credible (Mills, 1997). Foucault wants to understand how particular conceptions of the world become fixed and pass as truth, and argues that through understanding the rules and relations that govern our discourses, we can see their arbitrary nature and begin to question them.

Whereas the arbitrariness of our knowledges may be a matter of anxiety and despair for some people, "Foucault sees it as a reason to be optimistic. Ideas and practices which have oppressive and unjust effects on people and limiting effects within knowledge and science can always be changed" (O'Farrell, 2006, p. 189). Foucault's goal with his archaeological projects is thus emancipatory. By showing the contingent and arbitrary constraints (rules of formation) on our discourses, Foucault aims to question taken-for-granted truths and so open up possibilities for new or oppressed discourses, knowledges, and truths. By showing that our knowledges are the discursive effects of certain historical events and practices (conditions of possibility), he shows that our discourses are not set in stone, that these can be changed. Archaeology thus enables "one to think differently about the present, to interrogate that in our contemporary experience which we take for granted, through an examination of the conditions under which our current forms of truth have been made possible" (Rose, N. S., 1996, p. 107).

In short, with his archaeological projects, Foucault aims to questions the body of knowledge that people would accept as self-evidently true. Archaeology aims to question and understand "on what type of assumptions, of familiar notions, of established, unexamined ways of thinking accepted practices are based" (Foucault, Michel 2001, p. 456). His archaeological projects ask why things are (understood in) the way that they are (understood)? Why are they so, and not otherwise? What discourses or discursive practices, what implicit 'rules', or what conditions, give rise to these current arrangements? In doing so, Foucault hopes to open up possibilities for other lines of thought, and other political and ethical possibilities that come with it.

4.2 The archaeological level of description: discourses as monuments

In the introduction of 'The Archaeology of Knowledge', originally 'L'archéologie du Savoir' from 1969, Foucault (2002a) explains that the aim of this book is to provide some clarification on the 'method' he has used in his earlier books 'Madness and Civilization', 'The Birth of the Clinic', and 'The Order of Things'. These books were never intended to be part of a single approach, Foucault only realised the consistency of his method after finishing 'The Order of Things' in 1966 (Gutting, 1989). Based on these earlier studies, in 'The Archaeology of Knowledge', Foucault tries to provide some reflection on his 'method'—rather than providing an explicit or instructive method, we will see that Foucault presents a particular level of description:

> —in so far as my aim is to define a method of historical analysis freed from the anthropological theme, it is clear that the theory that I am about to outline has a

> dual relation with the previous studies. It is an attempt to formulate, in general terms … the tools that these studies have used or forged themselves in the course of their work. But, on the other hand, it uses the results already obtained to define a method of analysis purged of all anthropologism. The ground on which it rests is the one that it has itself discovered. (Foucault, 2002a, p. 17)

As also discussed in the previous chapter, Foucault does not want to appeal to the human subject and its consciousness as the source of all meaning. Next to eliminating the central role of the constituting subject, with archaeology he also does not want to search for the 'causes' of discourse, such as the Marxist search for the material determinants and social factors of ideology. Neither does he want to search for the deeper meaning of practices and discourse. Where hermeneutics tries to interpret discourse to recover the meaning behind the statements and reconstruct the thoughts of the authors, Foucault aims to take statements as objects of study in their own right:

> Archaeology tries to define not the thoughts, representations, images, themes, preoccupations that are concealed or revealed in discourses; but those discourses themselves, those discourses as practices obeying certain rules. It does not treat discourse as *document*, as a sign of something else, as an element that ought to be transparent, but whose unfortunate opacity must often be pierced if one is to reach at last the depth of the essential in the place in which it is held in reserve; it is concerned with discourse in its own volume, as a *monument*. It is not an interpretative discipline: it does not seek another, better-hidden discourse. It refuses to be 'allegorical'. (Foucault, 2002a, p. 155)

This characteristic principle—to treat discourse as 'monuments' rather than 'documents'—explains Foucault's choice for the archaeological metaphor.

> There was a time when archaeology, as a discipline devoted to silent monuments, inert traces, objects without context, and things left by the past, aspired to the condition of history, and attained meaning only through the restitution of a historical discourse; it might be said, to play on words a little, that in our time history aspires to the condition of archaeology, to the *intrinsic description of the monument*. (Foucault, 2002a, p. 8; italics are mine)

In his archaeological approach, Foucault is thus not interested in studying the meaning of discourses and their objects, as ‘documents’. Rather, by treating discourses as meaningless objects, as ‘monuments’, Foucault hopes to show that discourses have their own, intrinsic, systematic order. By describing the ‘discursive regularities’ at the level of discourse itself (see previous chapter), Foucault aims at the ‘intrinsic description of the monument’—the archaeological level of description.

Further in line with this archaeological metaphor, Foucault wants to focus on the overall configuration of the ‘site’ from which an object is excavated (Gutting, 2005). Together with his explicit rejection of humanist approaches, this understanding of archaeological analysis has occasionally resulted in Foucault being labelled a ‘structuralist’, which he irately rejects: “certain half-witted ‘commentators’ persist in labelling me a ‘structuralist’. I have been unable to get it into their tiny minds that I have used none of the methods, concepts, or key terms that characterize structural analysis” (Foucault, 2002b, p. xv).

Structuralism argues that individual components of a system do not have any intrinsic meaning, but only have meaning in relation to the structure as a whole. Saussure, who is generally seen as the father of structuralism, explains in his seminal book ‘Theory of Language’ (1974), that languages contain differences and relationships, which constitute the identity and meaning of the individual elements. It is the structure itself that gives meaning to its components; components gain meaning by means of relations and differences to other components. In explaining the creation of meaning this way, structuralism explicitly steps away from approaches that take meaning to originate in human experience.

Despite Foucault’s own objections, some of the main characteristics of structuralism can be found in archaeology as well. As such, many scholars label Foucault’s work as ‘post-structuralist’ (Ball, 1995; Graham, 2011; Humes & Bryce, 2003; Wetherell, Yates, & Taylor, 2001; Willig, 2008; Zeeman, Poggenpoel, Myburgh, & Van der Linde, 2002). Whereas Foucault and post-structuralism both operate with the ‘linguistic turn’—which sees language as being constitutive; it produces both the (social) world and the subject—other approaches concerned with discourse and language, such hermeneutics and structuralism can be said to function under a different condition of possibility, that of the modern episteme or the logic of representation. Having said this, in this thesis I am not interested in labelling Foucault's work or his epistemological stance(s). This would constitute a rather different project than my more practically oriented attempt to bring Foucault's archaeological approach to the applied disciplines of human factors and safety science. As such, in outlining Foucault's archaeological approach, here it suffices to say that both (post)structuralism and Foucault are looking for the 'structures' within discourse. However, where structuralism seeks for atemporal structures, Foucault's (2002a) archaeological approach seeks to describe historically contingent rules and relations that characterise a discursive practice. As Foucault forcefully explains, the regularities that he searches for in his archaeological projects,

> … [do] not constitute, above events, and in an unmoving heaven, an atemporal structure; it is defined as the group of rules that characterize a discursive practice: but these rules are not imposed from the outside on the elements that they relate together; they are caught up in the very things that they connect; and if they are not modified with the least of them, they modify them, and are transformed with them into certain decisive thresholds. (2002a, p. 144)

Foucault (2002a) argues that every mode of thinking involves some (internal) 'group of rules', besides the formal constraints of grammar and a logical narrative, that deem certain groups of statements as meaningful and others as nonsensical. "Discourse is constituted by the difference between what one could say correctly at one period (under the rules of grammar and logic) and what is actually said. The discursive field is, at a specific moment, the law of this difference" (Foucault, 1991b, p. 63). In his archaeological works, he looks into describing the thin subset of things that are actually said—the 'archive'—to reveal the historically contingent rules that further narrow the discursive options for what can be credibly said and thought at a particular time and place. In the introduction of 'The Order of Things', Foucault wonders:

> … whether there do not exist, outside their customary boundaries, systems of regularities that have a decisive role in the history of sciences. I should like to know whether the subjects responsible for scientific discourse are not determined in their situation, their function, their perceptive capacity, and their practical possibilities by conditions that dominate and even overwhelm them. In short, I tried to explore scientific discourse not from the point of view of the individuals who are speaking, nor from the point of view of the formal structures

of what they are saying, but from the point of the rules that come into play in the very existence of such discourse … (Foucault, 2002b, pp. xiv-xv)

In stepping away from the constitutive subject towards searching for 'systems of regularities' and knowledge that creates its own 'conditions of existence', Foucault explicitly differentiates between the "epistemological level of knowledge" and the "archaeological level of knowledge" (Foucault, 1981, p. xiv). As an example, he explains that in the Classical period:

> … unknown to themselves, the naturalists, economists, and grammarians employed the same rules to define the objects proper to their own study, to form their concepts, to build their theories. It is these rules of formation, which were never formulated in their own right, but are to be found only in widely differing theories, concepts, and objects of study, that I have tried to reveal, by isolating, as their specific locus, a level that I have called, archaeological. (Foucault, 2002b, p. xii)

In summary, with his archaeological approach, Foucault aims to provide a "project of a *pure description of discursive events*" (Foucault, 2002a, p. 29, italics in original). By treating discourses as monuments, rather than documents, the archaeological level of description moves away from the constituting epistemic subject towards an emphasis on describing the systematic order intrinsic to discourse. As such, archaeology can be seen as level of description, rather than a specific or instrumental methodological approach. In the following sections, I aim to describe some specific analytical concepts through which Foucault hopes to accomplish such an archaeological description.

4.3 The regularities and unities of discourse: systems of formation

So far we have seen that Foucault's archaeological approach aims to describe the historically contingent 'discursive regularities' or "discourses as practices obeying certain rules" (Foucault, 2002a, p. 155) that dictate what can be said and thought at a particular time and place. In further elaborating on his archaeological approach, Foucault explains that he aims to describe the "relations that may legitimately be described between … statements" (Foucault, 2002a, p. 34). He wonders what makes certain groups of statements into coherent 'individualised' discourses. "What, in fact, are medicine, grammar, or political economy?" (p. 35); "what are these unities?" (p.34) "Are they merely a retrospective regrouping by which the contemporary sciences deceive themselves as to their own past? Are they forms that have become established once and for all and have gone on developing through time" (Foucault, 2002a, p. 35)? Foucault calls such 'unified' discourses, 'discursive formations'. A discursive formation may be the body of knowledge of a particular scientific discipline, however, Foucault emphasizes that discursive formations are not always equivalent to scientific disciplines.

In investigating the unities of discourse, Foucault concludes that the unity of a discursive formation does not stem from the unity provided by any of its elements—its objects, forms of statement, concepts, or themes. For example, he "realized that the unity of the object 'madness' does not enable one to individualize a group of statements, and to establish between them a relation that is both constant and describable" (Foucault, 2002a, p. 35). He realises that it is none of these elements of discourse, as a single discursive formation does not need to refer to a single object (or system of objects), will have a variety of 'enunciative modalities' (places of speaking from which statements are enunciated), uses changing and new conceptual frameworks, and may develop diverse theoretical viewpoints. Rather, Foucault recognises that it is the regularity in the formation of these discursive elements that constitutes the unity of discursive formations: "if there really is a unity, it does not lie in the visible, horizontal coherence of the elements formed; it resides, well anterior to their formulation, in the system that makes possible and governs that formulation" (Foucault, 2002a, p. 80). For Foucault, discursive formations spring from a unifying 'system of dispersion', which defines a 'field' within which a variety of different, possibly conflicting, sets of elements can be deployed (Foucault, 2002a; Gutting, 1989).

> Whenever one can describe, between a number of statements, such a system of dispersion, whenever, between objects, types of statement, concepts, or thematic choices, one can define a regularity (an order, correlations, positions and functioning, transformations), we will say, for the sake of convenience, that we are dealing with a *discursive formation* … The conditions to which the elements of this division (objects, mode of statement, concepts, thematic choices) are subjected we shall call the *rules of formation*. The rules of formation are conditions of existence (but also of coexistence, maintenance, modification, and

> disappearance) in a given division. (Foucault, 2002a, pp. 41-42; italics in original)

As the 'space' where these statements emerge, the system of dispersion can be described by the 'rules of formation' that govern the formulation of these various elements. "This dispersion itself ... can be described in its uniqueness if one is able to determine the specific rules in accordance with which its objects, statements, concepts, and theoretical options have been formed" (Foucault, 2002a, p. 80). These 'rules of formation' are at the heart of Foucault's archaeological analysis of discourse. As such, 'The Archaeology of Knowledge' focuses on specifying these rules of formation by breaking them down in four categories, the rules of formation of objects, enunciative modalities, concepts, and the formation of strategies.

After sketching his elaborate approach for analysing the formation of the four elements that make up discursive formations—the next section will provide Foucault's suggestions for describing the rules of formation of objects—Foucault questions whether his approach to studying discursive formations really points to unities: if it is capable of 'individualising' wholes. He finds that the different levels of formation are not independent of each other. In explaining the connections among the various rules of formation that govern discursive formations, he points to 'a vertical system of dependencies' or 'hierarchy of relations' that exist between the formation of the various discursive elements. As an effect of these relations, "not all the positions of the subject, all the types of coexistence between statements, all the discursive strategies, are equally possible, but only those authorized by anterior levels" (Foucault, 2002a, p. 81).

Moreover, Foucault sees the unity of systems of formation in their mobility, in their ability to adapt over time. He does not see discursive formations as originating from some initial point— "beginning, origin, foundation, system of axioms" (Foucault, 2002a, p. 83)—on the basis of which history has merely to unfold in a necessary way. For him, systems of formation are not a number of static rules that, once formed, define the possibilities and characteristics of a discursive formation once and for all. Rather, he explains by example, even though in the nineteenth century, "criminal jurisprudence, demographic pressure, the demand for labour, the forms public assistance, the status and juridical conditions of internment, were continually changing," the discursive formation of psychiatry "continued to establish the same group of relations between these elements; in this way, the system preserved the characteristics of its individuality" (Foucault, 2002a, p. 83).

Within the same rules of formation, new objects can appear, new enunciative modalities emerge, new concepts are outlined, and new theoretical structures can be built. "A discursive formation, then, does not play the role of a figure that arrests time and freezes it for decades or centuries; it determines a regularity proper to temporal processes" (Foucault, 2002a, p. 83). Moreover, the rules of formation might change over time, and, inversely, the discursive practices modify the domains that they relate to in their turn. For example, the hospital "did not remain unaffected when clinical discourse was put into relation with the laboratory: the body of rules that governed its working, the status accorded the hospital doctor, the function of his observation … , were necessarily modified" (Foucault, 2002a, p. 84). In stronger words: "Discourse and system produce each other" (Foucault, 2002a, p. 84).

In eloquently summarising his intent and providing some reflections on his approach to capture discursive formations as 'systems of formation', Foucault explains that the rules of formation that he has been talking about do not exist prior to the discourse that constitutes them. The regularity of statements is defined by the discursive formation itself. "The fact of its belonging to a discursive formation and the laws that govern it are one and the same thing" (Foucault, 2002a, p. 131). He continues, "this is not paradoxical since the discursive formation is characterized not by principles of construction but by a dispersion of fact…" (Foucault, 2002a, p. 131). In summary:

> These systems of formation must not be taken as blocks of immobility, static forms that are imposed on discourse from the outside, and that define once and for all its characteristics and possibilities. They are not constraints whose origin is to be found in the thoughts of men, or in the play of their representations; but nor are they determinations which, formed at the level of institutions, or social or economic relations, transcribe themselves by force on the surface of discourses. These systems – I repeat – reside in discourse itself; or rather (since we are concerned not with its interiority and what it may contain, but with its specific existence and with its conditions) on its frontier, at that limit at which the specific rules that enable it to exist as such are defined. By system of formation, then, I mean a complex group of relations that function as a rule: it

> lays down what must be related, in a particular discursive practice, for such and such an enunciation to be made, for such and such a concept to be used, for such and such a strategy to be organized. To define a system of formation in its specific individuality is therefore to characterize a discourse or a group of statements by the regularity of a practice. (Foucault, 2002a, p. 82)

4.4 The rules of formation of objects & discursive relations

In The Archaeology of Knowledge, in meticulously outlining the rules of formation for these various discursive elements, Foucault admits that his three earlier archaeological studies—'Madness and Civilization', 'Birth of a Clinic', and the 'Order of the Things'—have not achieved a complete description of dispersion of the four elements of discursive formations (as outlined in the various section). He explains that for these various discursive domains he intended to describe "the rules for the formation of objects, modalities of statement, concepts, and theoretical choices. But it turned out that the difficult point of the analysis, and the one that demanded greatest attention, was not the same in each case" (Foucault, 2002a, p. 72). He explains that in the discursive formation that he deals with in 'Madness and Civilization', psychopathology, "the problem lay in the emergence of a whole group of highly complex, interwoven objects; it was necessary above all to describe the formation of these objects, in order to locate in its specificity the whole of psychiatric discourse" (Foucault, 2002a, p. 72). In "*Naissance de la Clinique*, the essential point of the research was the way in which, at the end of the eighteenth and the beginning of nineteenth century, the enunciative forms of medical discourse had been modified" (2002a, p. 72) and that "in *The Order of Things*, my attention was concentrated mainly on the networks of concepts and their rules of formation … as they could be located in General Grammar, Natural History, and the Analysis of Wealth".

The specific 'focus' on only one specific element of discourse in each of Foucault's earlier books is probably the largest discrepancy between Foucault's earlier archaeologies and the general method he proposes in 'The Archaeology of Knowledge'. Despite Foucault's explicit claim that the different levels of formation—that of objects, enunciative modalities, concepts, and theories—are not independent of each other (see previous section), I will select from Foucault's comprehensive approach just the facets that focus on the discursive formation of objects. Given this thesis' aim to analyse some objects from the human factors and safety science discourse, I will thus be inspired by Foucault's archaeological approach in 'Madness and Civilization', in which he deals with a similar task of studying the formation of objects.

In 'Madness and Civilization' (originally 'Folie et déraison' from 1961), Foucault (2001b) studies the formation of objects within the discursive formation of psychopathology, which was trying to define criteria for recognising and labelling the mad. In elaborating on the sudden emergence, or 'manifestation', of the object of madness, Foucault explains:

> There must have formed, silently and doubtless over the course of many years, a social sensibility, common to European culture, that suddenly began to manifest itself in the second half of the seventeenth century; it was this sensibility that suddenly isolated the category destined to populate the places of confinement. (Foucault, Michel, 2001b, p. 42)

In providing methodological instructions for studying the 'formation of objects' in 'The Archaeology of Knowledge', Foucault (Foucault, 2002a) draws heavily on examples from the discourse of psychopathology from the nineteenth century.

> The objects with which psychopathology has dealt … are very numerous, mostly very new, but also very precarious, subject to change and, in some cases, to rapid disappearance: in addition to motor disturbances, hallucinations, and speech disorders (which were already regarded as manifestations of madness, although they were recognized, delimited, described, and analysed in a different way), objects appeared that belonged to hitherto unused registers: minor behavioural disorders, sexual aberrations and disturbances, the phenomena of suggestion and hypnosis, lesions of the central nervous system, deficiencies of intellectual or motor adaptation, criminality. And on the basis of each of these registers a variety of objects were named, circumstances scribed, analysed, then rectified, redefined, challenged, erased. Is it possible to lay down the rule to which their appearance was subject? (pp. 44-45)

In studying the formation of such objects, Foucault asks: "What has ruled their existence as objects of discourse". In searching for answers, he provides elaborate instructions to describe the ‘rules for the formation of objects’, which he divides into three types, the ‘surfaces of emergence’, ‘authorities of delimitation’, and ‘grids of specification’.

“(a) First we must map the first *surfaces* of their *emergence*” (Foucault, 2002a, p. 45; italics in original). The first rule for the formation of objects thus asks us to describe the ‘surfaces of emergences’, the social sphere where the objects of a particular discursive formation emerge. He explains that any society’s social norms turn certain practices, individual differences, or behaviours, into objects of (scientific) investigation and concern, and therefore an object in the domain of a particular discursive formation (Foucault, 2002a; Howarth, 2000). For example, if a child’s behaviour is sufficiently deviant from the social norms (for example from those of the family), the child is considered to be mentally ill and becomes an object of concern for psychopathology. The family is then *a* surface of emergence for objects of the discursive formation of psychopathology (Gutting, 1989). Foucault explains about the surfaces of emergence for the object of madness: “In the case of nineteenth-century psychopathology, they were probably constituted by the family, the immediate social group, the work situation, the religious community”, which he argues: “are all normative, which are all susceptible to deviation, which all have a margin of tolerance and a threshold beyond which exclusion is demanded” (Foucault, 2002a, p. 45).

Secondly, Foucault’s approach asks: “(b) We must also describe the authorities of delimitation” (Foucault, 2002a, p. 46). The ‘authorities of delimitation’ refer to those to whom a society gives the authority to determine which objects belong to which discursive formation. For example,

> … in the nineteenth-century, medicine (as an institution possessing its own rules, as a group of individuals constituting the medical profession, as a body of knowledge and practice, an authority recognized by public opinion, the law, and government) became the major authority in society that delimited, designated, named, and established madness as an object. (Foucault, 2002a, p. 46)

However, Foucault explains, medicine was not alone in this, as the law (penal law in particular), the religious authority, as well as literary and art criticism, also ‘delimited, designated, named, and established’ madness as an object. The question thus becomes what objects belong to what discursive formation—and who decides, who has the authority to allocate these objects—as this has severe consequences for the meaning, function, and effects of these objects.

As a third type of rule for the formation of objects, Foucault proposes that: "(c) Lastly, we must analyse the *grids of specification*" (Foucault, 2002a, p. 46; italics in original), the systems through which discursive formations classify and relate various kinds of objects. In Foucault's own words, with the 'grid of specification' he means "the systems according to which the different 'kinds of madness' are divided, contrasted, related, regrouped, classified, derived from one another as objects of psychiatric discourse" (Foucault, 2002a, p. 46). He explains that for nineteenth-century psychopathology the grids of differentiation were the soul, conceived as a specific system of hierarchized, interrelated faculties; the body, conceived as a system of interdependent organs; the life and history of individuals; and the interplays of neuropsychological correlations. "Thus, someone with a certain sort of chemical imbalance or inherited disposition might, by that very fact, be an object of psychopathology" (Gutting, 1989, p. 235). Grids of specification thus function to classify and relate various sorts of objects on the basis of the properties that they posses or the symptoms they exhibit (Howarth, 2000).

In short, Foucault thus asks us to describe how objects emerged and where they came from, who provides them with their legitimacy, and what their concrete operation and content is. However, Foucault argues that such descriptions in themselves are inadequate, as these planes of emergence, authorities of delimitation, or forms of specification do not provide objects, "fully formed and armed, that the discourse of psychopathology has then merely to list, classify name, select, and cover with a network of words and sentences" (Foucault, 2002a, p. 47). He explains that we should not see discourse as a place where previously established objects emerge: "it is not easy to say something new; it is not enough for us to open our eyes, to pay attention, or to be aware, for new objects suddenly to light up and emerge out of the ground" (Foucault, 2002a, p. 49). We should not see objects as 'discoveries' that can readily be taken up by any willing (scientific) discourse. "Such facts lie beyond the grasp of contemporary research: indeed the problem is how to decide what made them possible, and how these 'discoveries' could lead to others that took them up" (Foucault, 2002a, p. 48).

The descriptions of surfaces of emergence, authorities of delimitation, and grids of specification have located several 'planes of differentiation' in which objects of discourse may appear, however, the question as to how these objects are formed and consequently taken up by particular discourses remains unanswered. Foucault finds the solution to the formation of objects in looking for a complex group of historically contingent relations. These 'discursive relations', as Foucault calls them, are relations between institutions, economic conditions, social processes, behavioural patterns, norms, types of classification, etc. (Foucault, 2002b). However, such 'discursive relations', are not internal to discourse—they do not establish a discursive organisation of logical propositions or sentences—nor are they external to discourse—they do not impose, from the outside certain forms or restrictions on it. Rather, Foucault explains:

> They are, in a sense, at the limit of discourse: they offer objects of which it can speak, or rather (for this image of offering presupposes that objects are formed independently of discourse), they determine the group of relations that discourse must establish in order to speak of this or that object, in order to deal with them, name them, classify them, explain them, etc. These relations characterize not the language (*langue*) used by discourse, nor the circumstances in which it is deployed, but discourse itself as a practice. (Foucault, 2002a, pp. 50-51).

Foucault thus argues that in understanding the formation of objects of discourse we need to describe the group of relations that characterise it as a discursive practice. In the case of the nineteenth-century psychiatric discourse he argues that the nineteenth-century psychiatric discourse "is characterized not by privileged objects, but by the way in which it forms objects that are in fact highly dispersed. This formation is made possible by a group of relations established between authorities of emergence, delimitation, and specification" (Foucault, 2002a, p. 49). Instead of referring to these as 'discursive relations', he occasionally refers to these as the 'conditions of possibility' for the emergence of certain knowledges (see next section).

Tying this back to Foucault's account of discursive formations, he explains that instead of seeking the unity of discourse in objects themselves, we are "sent back to a setting-up of relations that characterizes discursive practice itself; and what we discover is neither a configuration, nor a form, but a group of rules that are immanent in a practice, and define it in its specificity" (Foucault, 2002a, p. 51). What then emerges is a unity of another type, a system of formation that remains stable. “But let there be no misunderstanding,” Foucault (2002a, p. 52) stresses,

> … it is not the objects that remain constant, nor the domain that they form; it is not even their point of emergence or their mode of characterization; but the relation between the surfaces on which they appear, on which they can be delimited, on which they can be analysed and specified.

4.5 Conditions of possibility

In the previous section, we have seen that in order to understand the ‘space’ in which possible objects can emerge, we need to capture the ‘discursive relations’—established between the surfaces of emergence, authorities of delimitation, and grids of specification—that characterise it as an object. Foucault thus suggests to describe the formation of objects by “relating them to the body of rules that enable them to form as objects of a discourse and thus constitute the conditions of their historical appearance” (Foucault, 2002a, p. 53). More generally, Foucault explains about the rules of formation: “The conditions to which the elements of this division (objects, mode of statement, concepts, thematic choices) are subjected we shall call the *rules of formation*. The rules of formation are conditions of existence … in a given division” (Foucault, 2002a, pp. 41-42; italics in original). This is why Foucault’s archaeological approach is different from other approaches. It does not try to capture the ‘conditions of validity’, to assess whether a particular discourse is true or not, but rather tries the ‘conditions of existence’ or the ‘conditions of possibility’ for certain discourses or discursive element—such as objects—to emerge. These conditions are:

> The conditions necessary for the appearance of an object of discourse, the historical conditions required if one is to ‘say anything’ about it, and if several people are to say different things about it, the conditions necessary if it is to exist in relation to other objects, if it is to establish with them relations of resemblance, proximity, distance, difference, transformation – as we see, these conditions are many and imposing. (Foucault, 2002a, p. 49)

As such, Foucault sometimes—particularly in ‘The Order of Things’ (Foucault, 2002b)—describes these relations in Kantian terms, saying that he seeks the ‘conditions of possibility' for thought at a given time and place. However, as opposed to Kant's search for universal constraints on experience, for Foucault the conditions of possibility are historical and vary across times, places, and bodies of knowledge. He asks, for example, "what conditions did Linnaeus … have to fulfil, not to make his discourse coherent and true in general, but to give it, at the time when it was written and accepted, value and practical application as scientific discourse…" (Foucault, 1981, pp. xiv-xv)?

To emphasise the idea that our discourses are historical productions, Foucault introduces—what he himself calls a "barbarous term" (2002a, p. 143)—the notion of 'historical a priori'. "An a priori not of truths that might never be said, or really given to experience; but the a priori of a history that is given, since it is that of things actually said" (Foucault, 2002a, p. 143). As such, Foucault is not interested in a transcendental a priori, "not a condition of validity for judgements, but a condition of reality for statements. It is not a question of rediscovering what might legitimize an assertion, but of freeing the conditions of emergence of statements…" (Foucault, 2002a, p. 143). Foucault's archaeological analysis of the conditions of possibility for our discourses thus takes place at the level of the 'archive', which is the assortment of existing discourses at a particular time and place, the collection of 'things actually said'. "Archaeology describes discourses as practices specified in the element of the archive" (Foucault, 2002a, p. 148).

In short, by focusing on the conditions of possibility for certain knowledges, discourses, or objects to emerge, archaeology is "an inquiry whose aim is to rediscover on what basis knowledge and theory became possible; within what space of order knowledge was constituted…" (Foucault, 2002b, pp. xxi–xxii). As such, Foucault's archaeological approach asks not about the 'conditions of validity' but rather about the 'conditions of possibility' for certain knowledges:

> It is not a question of rediscovering what might legitimize an assertion, but of freeing the conditions of emergence of statements, the law of their coexistence with others, the specific form of their mode of being, the principles according to which they survive, become transformed, and disappear. (Foucault, 2002a, p. 143)

4.6 Conclusion

In the previous chapter we have seen that Foucault sees discourse as constitutive, as systematically creating the objects and subjects of knowledge. Discourses dictate what can be said and thought at a particular time and place. With his archaeological approach, Foucault aims to question taken-for-granted 'truths' and discourses. By outlining the assumptions and other unexamined ways of thinking our discursive practices are based, Foucault hopes to show that discourses are historically contingent and that they are thus open to scrutiny and change.

In his archaeological projects, Foucault argues that the constraints on discourse are not external determinations that are imposed on people's thoughts, but rather that the regularities of our discourses are internal to discourses themselves. As such, in his archaeological approach, Foucault is not interested in studying the meaning of discourses and their objects, as 'documents', but rather he treats them as meaningless objects, as 'monuments'. In doing so, Foucault hopes to describe the regularities of discourses in their own, intrinsic, systematic order. As such, archaeology can be seen as level of description, rather than a specific or instrumental methodological approach.

In the 'Archaeology of Knowledge', Foucault (Foucault, 2002a) explains that the (internal) unity of our discursive formations arises from the manner in which its elements—objects, enunciative modalities, concepts, or theoretic choices—are formed and dispersed. The system of formation can be described through the 'rules of formation' for these various elements. In focusing on the formation of objects, Foucault suggests to describe the 'surfaces of emergence', the 'authorities of delimitation', and the 'grids of specification' of objects, as well as the 'discursive relations' among these aspects. Another analytical concept that Foucault offers is that of the 'conditions of possibility' that characterise discursive practices. It is important to note that both the rules of formation nor the conditions of possibility have a transcendental status; that is, they are not "principles of construction" that exist prior to the discourse that constitutes them (Foucault, 2002a, p. 82). Nor do they form the internal constitution of the object. For Foucault there is nothing more than the totality of historically existing discourse—all captured in the 'archive'. The rules of formation for discursive formations (including its objects) as well as the conditions of possibility are descriptions of the relations and regularities of discourse itself.

With his archaeological approach, Foucault convincingly discards positivist, humanist, structuralist, hermeneutic, and essentialist accounts that reduce discourse to a pre-existing reality or see it emerge from a constituting subject. What Foucault is concerned with in his archaeological projects is to describe it its own complexity:

> To substitute for the enigmatic treasure of 'things' anterior to discourse, the regular formation of objects that emerge only in discourse. To define these *objects* without reference to the *ground*, the *foundation of things*, but by relating them to the body of rules that enable them to form as objects of a discourse and thus constitute the conditions of their historical appearance. To write a history of discursive objects that does not plunge them into the common depth of a primal soil, but deploys the nexus of regularities that govern their dispersion. (Foucault, 2002a, pp. 52-53)

In short, with his archaeological approach Foucault aims to analyse discursive formations and the rules and relations that define the conditions of possibility for thought at a particular time and place. As such, to stress the point, Foucault is not interested in the 'truth' of any particular discourse. He is chiefly interested in the describing the regularities he sees in discourses in order to show or question some of the unexamined ways of out thinking. This, eventually, will lead to questions about the effects—theoretically, practically, ethically—of our historically contingent discursive practices. This, however, is where Foucault's genealogy comes in (see chapter 8). I will further reflect on this in the conclusion of this thesis.

For now, it suffices to say that the analytical studies presented in this thesis, in trying to pursue Foucault's suggestions for the archaeological level of description, draw chiefly on a number of the analytical concepts for archaeological analysis as outlined in this chapter (and the previous). The two papers presented in Part III will focus respectively on the rules of formation and the conditions of possibility for situation awareness to emerge in the scientific discourse. The last two papers, presented in Part IV, after looking at how the object of their analysis is constituted in the discourse, will also start to ask questions about the effects of subjectivity that their respective discursive object has. The following chapter will provide a number of further elaborations of the methodological and analytical choices made in these studies.

Chapter 5:
Methodological reflections

Abstract

Archaeology presents a level of analysis, rather than a prescriptive method for analysing the discursive formation of objects. The four analytical studies in this thesis develop and present their own specific methodological approaches that are inspired by archaeology. As the review of literature is typically part of the archaeological process, the studies draw (to a greater or lesser extent) on the 'systematic literature review' as a transparent and methodical approach to select and analyse data. The primary sources of data that the studies use are peer-reviewed journal articles concerned with the objects of interest. In keeping with Foucault, these articles serve as the 'materialities' to base the archaeological analyses on, not as 'evidence' for making truth claims. Rather than reviewing the literature in order to essentialise it, or to find its true meaning, the archaeologically inspired studies describe the regularities that characterise the same objects that feature in the human factors and safety science discourses.

5.1 Archaeological inspiration

The previous chapters have intended to sketch and clarify Foucault's notion of discourse and his archaeological approach. They do so in depth, because these notions inspire the method that is applied in the various analytical studies that aim to describe the constitution of certain objects of the discourses of human factors and safety science—those of situation awareness, safety culture, and resilience. It is here, in these analytical studies, that the archaeological approach finds its final forms. As these chapters are written as publishable papers, rather than thesis chapters, they each present their specific methodological considerations themselves. This chapter, however, aims to provide some methodological deliberations that underlie the various archaeological approaches that the papers display.

First of all, it is important to state that the studies in my thesis are strongly influenced by Foucault's archaeological approach, however, in no simple manner would I dare to say that any of the studies, or the overall thesis, provides an archaeology. The thesis does not provide an archaeology of human factors as a discursive formation, neither does it so for safety science. It also does not provide an archaeology at the level of objects—whether situation awareness, safety culture, or resilience—as Foucault (2001b) presents in 'Madness and Civilisation'. As bodies of knowledge, the disciplines of human factors and safety science may not even provide very good objects for archaeological analysis. Who is to say that they present unified discourses? Until someone, in a larger archaeological project—which would take more than a PhD project—has established that the 'objects', 'enunciative modalities', 'concepts', and 'theoretic choices' of either of those disciplines emerge from a single 'system of dispersion' (see previous chapter), we should not even be referring to these disciplines as discursive formations. For this reason, in my thesis I consistently refer to human factors and safety science as (scientific) disciplines.

All that I claim for the four analytical studies in my thesis is that they are *inspired* by Foucault's suggestions for archaeological analysis. That is, they draw on one or more of Foucault's analytical concepts for the archaeological analysis—most prominently the 'conditions of possibility' and 'rules of formation' for the formation of objects (see the previous chapter). Also, rather than pursuing questions about the ontological or epistemological status of the objects of my analyses, the papers in my thesis try to study objects as historically contingent constructions that have certain effects. The studies are not interested in the practical utility of these objects, they aim not to be judgmental, but rather are concerned with the manner in which the objects of interest discursively function. As such, I would also dare to claim that the reasons for my various studies are in line with the emancipatory goals that Foucault pursues in his archaeological projects: questioning the taken-for-granted nature and 'truths' of dominant discourses.

Even though the various archaeologically inspired studies that I present in this thesis diverge from Foucault's (2002a) suggestions for a 'full' archaeology, as outlined in 'The Archaeology of Knowledge', Foucault encourages this kind of 'adapted' use of his work. He explains: "I would like my books to be a kind of tool-box which others can rummage through to find a tool which they can use however they wish in their own area..." (Foucault, 1974, translated by O'Farrell, 2006, p. 50). Inspired by Foucault's archaeological projects—primarily that of ‘Madness and Civilization’ (Foucault, Michel, 2001b), in which he studies the formation of objects, and his more general suggestions for archaeological analysis in ‘The Archaeology of Knowledge’ (Foucault, 2002a)—this is exactly the manner in which my various studies use archaeology: as ‘a tool’ for analysing the emergence and effects of some objects of the human factors and safety science discourses. As such, even though the four analytical studies are merely ‘archaeologically inspired’, throughout the thesis I refer to them as ‘archaeological studies’.

5.2 From prescribing method to rigorous scholarship

Apart from some ‘suggestions’ that Foucault presents in ‘The Archaeology of Knowledge’, archaeology does not present a ‘ready-made’ method for studying the discursive constitution of objects. Archaeology presents a ‘level of analysis’ rather than an explicit methodological approach. Foucault dislikes the idea of prescribing how things should be, or should (methodologically) be done: “I take care not to dictate how things should be” (Foucault, Michel, 2001a, p. 288). Rather, Foucault wants people to use his books as examples, as inspirations, as ‘tools’, “which they can use however they wish in their own area...” (Foucault, 1974, translated by O'Farrell, 2006, p. 50). As such, he explicitly refrains from providing a mechanical method on how to go about ‘doing’ archaeological analysis of discourse (or genealogy for that matter): “there exists no strictly Foucauldian method of analysing discourse” (Hook, 2001, p. 521).

Over the last decades, however, numerous studies and methods that analyse discourse from a Foucauldian perspective have been presented, proposed, and popularised: from studies presenting ‘Critical Discourse Analysis’ (CDA), which focuses mostly on the structural and linguistic features of texts, to various suggestions for ‘Foucauldian Discourse Analysis’ (FDA), which are often characterised by a genealogical focus on how discourse produces particular subjectivities and meanings about the social relationship between objects, people, and institutions (e.g. Powers, 2007; Waitt, 2005; Willig, 2008). The more Foucauldian approaches to analysing discourse, however, share Foucault’s reluctance to declare their method, fearful perhaps of being prescriptive. Graham explains that, “those using discourse analysis with Foucault shy away from prescribing method, for no matter how standardised the process, the analysis of language by different people will seldom yield the same result” (Graham, 2011, p. 667). Similarly, Arribas-Ayllon and Walkerdine (2008, p. 91) explain that, “there are no set rules or procedures for conducting Foucauldian-inspired analyses of discourse”.

This characteristic unwillingness to declare method has resulted in Foucauldian approaches being accused of presenting unsystematised speculation (Graham, 2005, 2011; Nixon & Power, 2007). Graham explains, for example, that “a perceived lack of precise methodological principles has lent weight to epistemological claims about the superior rigour of linguistically ‐ based methodologies, such as CDA, over those informed by the work of Foucault and other post-theorists” (2005, p. 4). It seems that in an attempt to avoid prescription, Foucauldian analysis of discourse becomes vulnerable to judgement against competing epistemological claims to methodological superiority.

> As such, it is important to stress—here in the method chapter of my thesis—that the lack of a standardised approach to analyse discourse does not equate to ‘unsystematic speculation’, nor is the lack of a prescribed method the same as being ‘imprecise’ or presents a ‘lack in rigour’ (Graham, 2011). We need to distinguish between the prescription of scientific method on the one hand and methodological rigour on the other.

As Foucauldian (archaeological) approaches to the analysis of discourse do not aim to establish a final account or reveal the true meaning of what is said, they do not require prescriptive or standardised methods. A standardised methodological approach allows for a study’s possible ‘replication’—for example to assess the ‘reliability’ of the initial study—it might increase the possibilities for ‘triangulation’—to assure the ‘validity’ of the analysis—and eventually it will help with the ‘generalisation’ of research findings into universal truth claims. However, as Foucault is not concerned with a positivist search for truth, there is no need for the systematisation and prescription of a methodological approach.

This, however, does not mean that one cannot develop methodological principles to analyse discourse or that one should not be academically rigorous in analysing discourse from a Foucauldian perspective. The commitment to a Foucauldian approach (or any other post-modern epistemology) does not relieve a researcher from being precise and transparent about his or her methodological and analytical choices. “It simply means that one has to be clear about objectives, limits and, most importantly, what one is doing" (Graham, 2011, p. 667). So, even though there is little agreement about how to achieve rigour in Foucauldian analyses of discourse (Bradbury-Jones, Irvine, & Sambrook, 2007; Nixon & Power, 2007), someone aiming to archaeologically analyse discourse, at the very least, needs to be explicit about what he or she is doing.

Instead of following a prescriptive method for analysing discourse, the various studies in my thesis develop their own methodological principles aimed at providing transparent and convincing narratives. Rather than presenting quantitative measures of the validity and reliability of my analyses[10], the studies aim to be upfront and transparent about the analytic choices and processes that constitute the analysis. As the studies are informed by and aim to be consistent with Foucault's suggestions for archaeological analysis, this means that, for the sake of clarity, they all explicitly declare and describe the archaeological concepts that are used for analysis. Also, by including extensive quotes form the various documents that are analysed, I hope to engage the reader in a critical analysis of the findings. Finally, in aiming to provide an 'authentic account' (Golden-Biddle & Locke, 1993), the results of the analyses are presented both in terms outcomes and (reflective) processes. This approach leaves critical readers with the possibility to (re)assess my analytic choices and attributions themselves.

5.3 Considerations of audience

As outlined in the introduction of my thesis, one of the main goals of pursuing these archaeologically inspired studies is to bring Foucault's explicitly anti-positivist—archaeological—approach to the practical and predominantly positivist disciplines of human factors and safety science. The thesis asks what insights an archaeological approach can provide in critically reflecting on some of the discursive objects that these disciplines are promoting. As such, the imagined audience of this thesis, or the four archaeological papers, is principally that of human factors and safety scholars. With this practically oriented audience in mind, two very important guidelines for both the thesis and the papers have been to 'keep it philosophically light' and to 'make it relevant'.

[10] Both terms—validity and reliability—are chief indicators of Western science's commitment to a positivist epistemology in which there is only one 'true' understanding of the world.

‘Light’ and ‘relevant’ are not the first two words that come to mind when describing Foucault’s archaeological projects. “Foucauldian theory is perceived as inaccessible and dangerous, which deters some researchers from engaging with this form of analysis, particularly those in more practice-oriented fields” (Graham, 2011, p. 664). Archaeology is an extremely elaborate and complex ‘method’, and it takes a Foucauldian purist (or philosopher) to appreciate its methodological delicacies and lengthy deliberations. Foucault’s genealogies, on the other hand, are generally easier to read and comprehend, and their practical ‘yield’ may be more evident. In bringing Foucault’s archaeological approach to the disciplines of human factors and safety science, I have had to make numerous ‘adjustments’—simplifications and analytical sacrifices—compared to Foucault’s ideas on discourse and archaeology as presented in chapters 3 and 4. These adjustments were necessary to make archaeology both appealing and acceptable for my intended audience.

Just one evident constraint that limits the extent to which the analytical papers are able to pursue archaeological arguments and descriptions, is the length of manuscripts that human factors and safety science journals are willing to accept for publication. Foucauldian (or sociological) journals understand the need for meticulous description within this genre and are typically happy to accept lengthy journal articles (‘Foucault Studies’ (2015) sets a target length of 12,000 words). The maximum number of words that is normally allowed by human factors or safety science journals, however, lies well under that of Foucauldian journals. Being set on writing to the human factors and safety science communities, I have had to rewrite many of my elaborate archaeological analyses into significantly shorter versions suitable for publication. These shorter versions, however, can be called less ‘archaeological’, as it was mostly the lengthy descriptions—aimed at describing ‘discourse as monuments’—and quotes—aimed at portraying ‘discourse in its own complexity’—that had to be sacrificed as a result of these strict word limits[11].

Considering the audience of my papers, another tactical choice in my papers has been to not dwell on providing various philosophical—epistemological—contemplations or even qualifications of my archaeological arguments. Instead of explicitly situating my

[11] The paper presented in chapter 6, which reviews the top hundred most cited papers on situation awareness, is still about 12,000 words in length after deleting many quotes and description. It remains to be seen if it can be published in this (lengthy) format.

archaeological claims in larger epistemological perspectives (such as (social) constructionism, constructivism, or post-structuralism), I have chosen to instead conclude the various papers with some rather specific conclusions that ask my audience to reflect on the effects of the 'constructions' that the papers provide. I feel that this 'tactical' approach to leave my audience with some 'food for thought' is not only the best way to translate some rather abstract ideas to a more practically oriented audience, but that it is also presents an appealing way to provide some modest reflections on the epistemological status of my claims.

Similarly, to make my archaeological arguments—that are typically rather 'dry' and theoretically heavy—more appealing, I have also made an effort to connect my archaeological findings to practical examples and relevant issues in the domains of human factors and safety science. In the chapter 7, which presents my paper on the conditions of possibility for situation awareness, for example, I try to connect my archaeological conclusions directly to the problematic case of Karl Lilgert, who is in jail for 'losing situation awareness'.

Despite the analytical sacrifices—both ideologically and practically—that come with writing to this practically oriented audience, I think that the analytic chapters preserve most of the archaeological principles as outlined in the previous two chapters. The simplifications and other necessary adjustments that are made are the necessary consequences of trying to build bridges between the competing goals and expectations of two diverging worlds: the practical 'real' world of human factors and safety science on the one hand, versus the carefully constructed anti-positivist world of Foucauldian archaeological analysis on the other.

5.4 Data

5.4.1 Archaeological data versus empirical evidence

Most academic disciplines expect a doctoral thesis to have an empirical component, either quantitative or qualitative. This thesis, however, is slightly different. In pursuing a Foucauldian—archaeological—approach, this is not a project for empirical data. An archaeological project, typically, does not involve any methods of observation or experimentation: it does not interview people, send out questionnaires, conduct some form of ‘in locus’ (ethnographic) field work, or pursue any other method for gathering ‘empirical’ data. However, in principle, various aspects of these methods can be used to capture discourse, which can subsequently be analysed at an archaeological level.

The difference, however, is that archaeology is not an empirical project in the simple sense that there is no ‘empiricism’. There is no ‘experience measurement’, because there is no place to stand ‘outside’ discourse in the external (empirical) world, collecting data that is (objectively) ‘given to the eyes’. In an archaeological project there is no ‘evidence’ either. As discussed in the previous chapters, with his conceptions of discourse and archaeology, Foucault wants to explicitly move away from a positivist approach to social analysis. It is not the goal of the archaeologist to produce a ‘true’ narrative, which needs to be supported by (empirical) evidence. Both words, ‘empirical’ and ‘evidence,’ are very far from any Foucauldian project.

This does not mean that there is no ‘data’ in archaeological projects. The data that Foucault draws upon in his archaeological projects are the ‘materialities’—the physical effects in the real—of the discursive practices that he is studying (Foucault, 1981). In his archaeological projects, Foucault focuses on studying discursive practices, whereas in his later, genealogical, projects he also starts to include non-discursive practices—such as the ways to organize time and space through architecture, etc. (see, for example, Foucault, 1991a). The ‘data’ for archaeological analysis are not just expressions of discourse, but the manifestation of discourse—the materialities of discursive practices as captured in the ‘archive’ (see previous chapter). It is, however, important to see the notion of archaeological data disconnected from the positivist approach of using it as evidence for truth claims. Just as the archaeologist studies discursive practices that produce the objects and subjects of which they speak, so is the discourse that the archaeologist writes productive. As Foucault explains provocatively, “I am fully aware that I have never written anything other than fictions” (Foucault, cited in Dreyfus & Rabinow, 1982, p. 204).

Foucault’s quote reinforces the idea, outlined in chapter 3, that instead of engaging in a battle of truth and fiction about the status of our knowledges, with his archaeological projects Foucault’s objective is to understand how particular discourses and objects might become formed and come to function with what might be the “effects of truth” (Foucault, 1980, p. 118). Reflecting on the status of archaeological research, this perspective means that all knowledge—including the (scientific) 'insights' produced by the researcher him or herself—is 'fiction'. The researcher thus does not study his data to 'discover' knowledge, but rather to 'author' it (Willig, 2008).

In short, in an archaeological project data is not used as (empirical) evidence that is objectively available to everybody—though observation or experimentation—to support (true) arguments, rather data is simply the discourse that the archaeologist draws upon to create his or her narrative, which is constitutive in its own right.

5.4.2 The data for my archaeological studies

This thesis is concerned with archaeologically studying situation awareness, safety culture, and resilience as objects of knowledge that emerge (primarily) in the discourses of human factors and safety science. These objects feature in various 'discursive modalities': books, articles, conferences, training courses, operating manuals, design guidelines, company guidelines, national policies, international laws or regulations, accident investigation reports, court transcripts, to name just a few. As objects are used in particular settings or configurations, with a particular intent and meaning, they are delimitated, specified, and reinforced as discursive objects. Every time that an object is invoked it is constituted and further legitimised as a meaningful discursive object. In my analyses, however, I focus on just the *scientific* production of these objects. The processes and rituals of science—particularly scientific publication—present a strong 'technology' for fixing a particular understanding of the world as 'truth' (Foucault, 1991a). As such, the data for the various archaeological studies in my thesis consists of the scientific part of the archive: the scientific discourses that invoke, and thereby constitute, the objects of situation awareness, safety culture, and resilience.

The materialities, in which scientific discourses primarily manifest, range from conference presentations and university tutorials to scientific books and journal articles. To limit my data set, I have particularly focused on peer reviewed journal articles as the materialities to study in my analyses. As the scientific peer-review process—the critical assessment by colleagues from the same field—is aimed at analysing and judging what is valid knowledge, it provides a good criteria for selecting discourses that are 'in the true' at this time and place (Foucault, 1981). Scientific books, book chapters, and conference proceedings are often not refereed, or at least to a lesser extent, and therefore do not reflect the same level of scientific acceptance as peer-reviewed journal articles.

The ‘scientific archive’—by which I refer to the “things actually said” (Foucault, 2002a, p. 143) in scientific discourses—is easily accessible nowadays. This is particularly so for the peer-reviewed scientific literature, as since the last decade almost all scientific journal articles are published in online equivalents and are easy to retrieve through various academic search engines and online citation databases. These digital resources make it easy to select and access scientific literature regarding a specific topic—object of knowledge—on the basis of various search criteria. The two most extensive academic citation databases, ‘Web of Science’ and ‘Scopus’, both have their own shortcomings. The biggest limitation, however, is that they both only provide the performance indicators, such as the number of times a paper has been cited, for the journals and citations that are in their respective databases (Durieux & Gevenois, 2010). As a number of recent studies point to Scopus as the preferred database for large-scale citation analysis of multidisciplinary scientific literatures (e.g. Adriaanse & Rensleigh, 2013; Aghaei Chadegani et al., 2013), Scopus (Elsevier, 2014) was chosen to access the ‘scientific archive’ and retrieve the materialities for the various studies in my thesis.

Using Scopus, the large amount of peer reviewed journal articles that are concerned with situation awareness, safety culture, or resilience were ranked and selected based on the number of times they have been cited in other scientific work. Because the number of citations for an journal article is regarded as an important indicator of scientific recognition and acceptance (Allik, 2013; Durieux & Gevenois, 2010; Schmoch & Schubert, 2008; Sharma, 2012), this final search and selection criterion provides a manner to select a part of the literature that has been widely accepted by the scientific community. When applicable, the specific search algorithms (used in Scopus) and the inclusion criteria for the respective studies are presented in the analytical chapters themselves.

It is important to note that the data sets for the various studies are neither exhaustive nor representative; they do not represent the entire ‘scientific archive’, all the things ever said, concerning these objects. There is no reference database that indexes all the papers that have been written about the objects of interest to the various studies. Moreover, the analytical choice to select papers on the basis of their number of citations makes the data set prone to numerous distortions: recent papers will be underrepresented, just as older papers that have been published before digital publication became standard. Also, because the number of citations for these papers is in constant flux, the group of top cited papers is an unstable data set. For example, just a few weeks after selecting the papers for the archaeological analysis of SA, the selection of the hundred most cited papers on SA would have made for a significantly different data set; some papers had more citations and would now be included in the data set, whereas some of the papers included in the initial data set would be omitted. However, just as archaeology is not concerned with using data as 'evidence' for making truth claims, neither is it concerned with data being representative of some larger cohort or truth.

The lack of requirement for 'representation' allowed the individual studies to move away from this systematic approach to selecting literature and to expand their initial data set by following up on seemingly relevant references, which were consequently added to the data set (see the next section about the systematic literature review as a methodological approach). This 'snowball' approach, in a few instances led the studies to include 'materialities' beyond just the scientific (peer-reviewed) literature. Where this is the case, these other materialities are clearly referenced for the sake of transparency.

In this section, I hope to have presented the analytical choices that make for a 'convincing' data set to base the narrative for the various archaeological analyses on. Simply put, the main data for the various studies in this thesis are the most cited peer-reviewed journal articles on situation awareness, safety culture, and resilience. These documents provide the (initial) materialities for analysis. In playing with words, but at the cost of being needlessly confusing, the selected materialities of the discourse could be said to serve as the 'empirical' material of my analytical studies.

5.5 Analysing data: archaeologically reviewing literature

The four analytical studies in this thesis each present their own specific archaeologically inspired analysis of their respective object of interest. As they develop their own way of analysing the data for their respective purposes, the specific methodological approaches and analytic choices are outlined in these papers themselves. However, in any archaeological project, the review of literature is typically part of the analytical process. As such, all four analytical studies 'review literature' on the basis of a number of analytical concepts—most prominently the 'conditions of possibility' and the 'rules of formation' for the formation of objects—to describe the constitution (and effects) of objects at an archaeological level. As such, in a broad understanding of the term, the four studies are conducted as 'literature reviews'. However, the notion of 'literature review' is an ambiguous term that comes with a number of connotations. In order to use this term to characterise the studies in this thesis, or, conversely, to set them apart from the literature review, the manner in which the studies 'utilise' the (systematic) literature review deserves some specification.

5.5.1 Archaeology versus the 'literature review'

To explain the manner in which my studies relate to and draw on the literature review, particularly that of the 'systematic literature review', I will rely on a strategy that Foucault often deploys: explaining what something is by explaining what it is not. A literature review is often motivated by the aim to summarise the essential or important points of the literature; it is about the 'essence' and 'content' of the literature. However, as we have seen in the previous chapter, archaeological studies are not interested in the essentialising discourse. Nor is literature reviewed from a hermeneutical perspective, with a focus on its (deeper) meaning. In no simple manner do the archaeological studies in this thesis aim to interpret the literature to understand the meaning of the objects of interest. They are not interested in understanding what the author 'really' means when referring to any of the objects of interest. In reviewing literature, the studies in this thesis also do not draw on any form of 'content analysis' as a formal technique or informal strategy for unearthing (in a Durkheimian manner) the true or essential meaning of documents. They will also not present a 'meta-analysis' (Glass, 1976) of the (essential) categories used in the literature.

The above listed aims for literature reviews are certainly relevant for their own purposes, however, archaeology does not aim to 'review literature' at that level. In an archaeological project, literature is not reviewed to essentialise it or to reduce the volume of the discourses in which the respective objects of interest emerge. Foucault explains (2002a, p. 53):

> When one describes the formation of the objects of a discourse, one tries to locate the relations that characterize a discursive practice, one determines neither a lexical organization, nor the scansions of a semantic field: one does not question the meaning given at a particular period to such words as 'melancholia' or 'madness without delirium', nor the opposition of content between 'psychosis' and 'neurosis'. Not, I repeat, that such analyses are regarded as illegitimate or impossible; but they are not relevant when we are trying to discover, for example, how criminality could become an object of medical expertise, or sexual deviation a possible object of psychiatric discourse. The analysis of lexical contents defines either the elements of meaning at the disposal of speaking subjects in a given period, or the semantic structure that appears on the surface of a discourse that has already been spoken; it does not concern discursive practice as a place in which a tangled plurality – at once superposed and incomplete – of objects is formed and deformed, appears and disappears.

It is thus the goal of the studies in this thesis that sets them apart from 'simply' being literature reviews. As opposed to analysing the content, meaning, or even truth, of the literature, my studies look to describe the discursive regularities and relations that characterise our current discursive practices, more specifically, the formation of certain objects. As such, the main focus of my studies is not on the content of the literature, but it is the literature—or better, the discursive practice—itself that is described. Archaeology takes the actual instances of discourse as the material 'data' to construct its 'review' on.

To bring home the point, archaeology differs from the traditional literature review, because archaeology is not primarily concerned with 'literature'. Archaeology is concerned with the materialities in which the discursive objects of interest emerge, rather than literature per se. Moreover, instead of 'reviewing' the contents of 'literature', archaeology aims to study discourse as monuments (rather than documents), 'viewing' them rather than 'reviewing' its contents. As such, despite the superficial similarities between archaeological studies and the traditional literature review, there are some caveats that need to be taken into account when labelling the studies in my thesis as 'literature reviews'.

Having said all of this, the (systematic) literature review does provide a number of useful—transparent and convincing, or at least 'accepted'—methodological guidelines for selecting and analysing the materialities for archaeological analysis. As such, despite the concerns outlined above, the studies in this thesis occasionally state to be conducted *as* a (systematic) literature review.

5.5.2 The systematic literature review as a tool for archaeological analysis

Despite the difference in objective and focus between the traditional literature review and my archaeological approach to reviewing literature (i.e. discursive practices), methodologically, the four archaeological studies are heavily influenced by the systematic literature review method. Two papers—'the emergence of SA' (chapter 6) and 'the rationale behind resilience' (chapter 9)—explicitly draw on the 'systematic literature review' as their methodological approach. The other two studies do so in a less explicit manner, but are nonetheless conducted in a similar manner—just less explicitly so. I call this a 'methodological' approach because the principles of the systematic literature review present a transparent strategy for the processes of selecting and analysing data, which are clearly issues of method.

Systematic literature reviews have been used to examine a wide range of contemporary and often contentious issues. "Complex 'real world' issues are not beyond the remit of systematic reviews" (Petticrew, 2001, p. 100). The systematic literature review opposes the more traditional 'narrative review', in which a carefully constructed narrative is argued on the basis of as much literature as the researcher considers relevant (Pickering & Byrne, 2013). Narrative reviews typically leave the sources, search strategies, and inclusion criteria for the review unspecified. Systematic literature reviews, on the other hand, are characterised by an explicit research approach: the sources and search strategies for literature are made explicit and the criteria for selection and analysis of the studies are uniformly applied. Synthesis often involves statistical summary. Through its systematic—methodical—and explicit approach about the choices made in the review, the systematic literature review aims to be transparent and reproducible (Collins & Fauser, 2004; Jones & Evans, 2000; Petticrew, 2001; Pickering & Byrne, 2013).

Typically, the analysis in the systematic literature review is conducted as a 'meta-analysis'—a term coined by Glass to depict the quantitative "analysis of analyses" (Glass, 1976, p. 3). The quantitative integration of data from the studies included in the data set is seen as the most efficient means of synthesising large volumes of papers (Pickering & Byrne, 2013). However, statistical synthesis is not a necessary component of a systematic literature review: "the lack of meta–analysis within a systematic review does not diminish its potential value" (Jones & Evans, 2000, p. 67). Qualitative analysis in the systematic review process is thus not without precedent and may be considered more appropriate for certain purposes (Petticrew, 2001). Particularly when Pickering and Byrne explain that the quantitative approach in a systematic literature review "is likely to prove effective for a wide range of disciplines that use a *positivist approach*" (2013, p. 9; italics are mine), it becomes evident that the assumption of positivism deems quantitative synthesis an inappropriate approach for this thesis' archaeological framework—which is explicitly anti-positivist. My studies can thus be said to be informed by what has been termed a 'qualitative systematic literature review' (Jones & Evans, 2000).

In line with the methodical approach of the (qualitative) systematic literature review, the two papers (chapter 6 and chapter 9) that explicitly draw on this approach select their data set (literature) based on pre-determined search and inclusion criteria. As discussed in the data section of this chapter, the specific algorithms and inclusion criteria will be presented in the studies themselves, however, here it is apt to discuss the general manner in which the data was used. After the selection, the literature was systematically—yet qualitatively—organised and analysed based on the various analytical concepts that Foucault (2002a) outlines for studying the 'formation of objects'. For chapter 6, 'the emergence of SA', this means systematically analysing the (almost) hundred papers in the data set on their 'surfaces of emergence' and 'authorities of delimitation', which will help to establish the 'discursive relations' that hope to shine some light on the archaeological questions of how and where SA emerges. For chapter 9, 'the rational behind resilience', this means dividing the analysis and qualitative description in the rationale or 'discursive need', the 'object', and 'subject' of resilience.

The systematic approach to selecting and analysing the data even allowed for some simple quantification in these papers. However, given the explicit anti-positivist approach of archaeology, it is of utmost importance to see these 'pie-charts' and ‘category counts’ for what they are, merely illustrations. To attach any additional value to them, is meaningless and might even undermine the archaeological point[12]. This, again, is where my studies move away from the systematic literature review. As we have seen, the data is in no simple way trying to be representative of a larger cohort of papers on the topic, nor does the analysis that precedes these simple forms of quantification claim to be definitive or even stable. There are no measures for the validity of these categorisations. There is no second, independent, researcher to establish the inter-rater reliability of any of the attributions. There is no statistical computation, no multivariate analysis of variance to assess the levels of significance, and thus no ‘95% confidence intervals’. Such ‘technologies’ (Foucault, 1991a) would project a false image of statistical ‘truth’ and distract from the qualitative approach in which these paper aim to pursue a description of the archaeological level of knowledge.

In short, all that these two studies—the two that are explicitly informed by the systematic literature review—take from the systematic literature review is its explicit and methodical approach for selecting and analysing papers; that is, they “use explicit methods of methodically search, critically appraise and synthesize the available literature on a specific issue” (Collins & Fauser, 2004).

The other two studies in this thesis—chapter 7, on the ‘conditions of possibility for SA’ and 8, which presents an ‘archae-genealogical study on the emergence and effects of safety culture as a scientific object’—adhere less strictly and explicitly to a systematic manner of selecting and analysing their respective data sets. The data—materialities—that these other two studies draw on is not as explicitly selected and thus as clearly delimitated. Initially, these studies also started with specific searches using online citation databases—searching the scientific archive for peer-reviewed papers on ‘situation(al) awareness’ and ‘safety culture’. However, in going though the initial set of papers that came out of these queries, other (seemingly relevant) references were pursued and included in the data set, sometimes even beyond the scientific (peer-reviewed) literature. This ‘snowball’ approach, in which the results of the initial search are used to retrieve and include further relevant papers, does however quickly grow into a lawless mass of papers, which opposes the transparent and methodical approach to selecting and including data that the systematic literature review promotes.

[12] To be honest, these ‘illustrations’ chiefly function adhere to the preference that most human factors and safety science journals have for the graphical summary of data.

Even though these two studies do not provide a similar systematic manner of selecting their data, the analysis of this 'amorphous' mass of data is systematically informed by Foucauldian concepts for analysis; all the literature was 'reviewed' (see previous section) on the basis of a selective set of analytical concepts. Chapter 7, 'the conditions of possibility for SA', tries to describe the regularities in the (entire) scientific body of literature on SA in terms of its 'conditions of possibility'. Chapter 8, 'the archae-genealogical study of safety culture' is the most dense and complex study of the four, as it aims to analyse the (unspecified) body of literature regarding the object of safety culture on a large number of Foucauldian concepts. Beyond the analytical concepts for archaeological inquiry—the paper aims to map the 'conditions of emergence', 'authorities of specification', and 'grids of specification' for safety culture as a scientific object—this study also analyses its data on a number of genealogical concepts: the 'disciplinary' and 'biopolitical' effects the object has, as well the as the form of 'governmentality' it constitutes.

The lack of an explicit methodological approach, beyond the underspecified explanations of 'archaeology', makes that these latter two studies read more as (narrative) literature reviews rather than 'analytical studies' (informed by the systematic literature review method). However, in distancing these studies from this characterisation, they 'review' literature at a different level and with a different aim: to produce archaeological arguments (see previous section, and chapter 4).

In short, even though the four studies differ in their commitment to a systematic approach, particularly in selecting relevant materialities, they all utilise part of the 'systematic literature review' method as part of the analytical process of (archaeologically) describing the constitution of objects in the discourse.

5.6 Conclusion

As archaeologies—or better, archaeologically inspired analyses—rather than literature reviews, the studies that follow in Part III and Part IV aim to describe the discursive practices and regularities that characterise the human factors and safety science discursive practices; in specific, those that constitute the objects of situation awareness, safety culture, and resilience. These chapters are all written as independent—stand-alone—journal articles, and as such, they all provide their own specific methodological assumptions and analytic choices. It is in these analytical studies, that the Foucault's archaeological approach finds its final forms.

It is, however, important to realise that there are numerous other ways in which Foucault's suggestions for archaeological analysis could have been 'operationalised' in the specific studies that aim to understand the discursive formation of objects. In Part III of this thesis, I present just two studies[13] that each draw on different, but central archaeological concepts for studying situation awareness as a discursive object: the rules of formation and conditions of possibility for the emergence of objects. However, I could also have chosen to analyse my data on the basis of other archaeological concepts. I could have tried to map the 'system of dispersion' from which situation awareness springs; I could have focused on just describing the 'discursive relations' as a complex web from which the object emerges. Similarly, Part IV presents two studies, focusing on the objects of safety culture and resilience, that both display only one possible archaeological narrative. Again, the objects could have been studied differently; on the basis of other archaeological or genealogical concepts and questions. As such, the four studies that will be presented in the following four chapters are just as contingent as the objects that they aim to study (I will provide further reflections on this in the conclusion of this thesis). Having said this, this does not take away from the intent of this thesis, which is to 'introduce' archaeological thinking, its framework, some of its analytical concepts, and level of analysis to the disciplines of human factors and safety science.

[13] A third paper, that asks about the 'grids of specification' for SA, is in the making, but as the paper does not add a new (archaeological) approach to study the object—it draws on the same study as presented in chapter 6—for the sake of brevity, this paper has been excluded from the thesis.

PART III

-

THE ARCHAEOLOGICAL ANALYSIS OF SITUATION AWARENESS

Chapter 6:
Situation awareness: The emergence of an object of knowledge

Summary

This paper sets out to explore the emergence of Situation Awareness (SA) as a scientific object. Foucault sees discourses as emerging under historically contingent 'rules of formation'. By examining the most cited journal articles on SA, this study finds that the 'surfaces of emergence' of SA can be characterised by high levels of complexity, systems control tasks, and a central concern for safety. The 'authorities of delimitation' for SA are characterised by their focus on studying either the human or the technological part of our socio-technological systems. These domains and disciplines draw on SA to deal with a number of performance problems that accompany the introduction of automation: information overload, cognitive workload, and the distancing of the operator from real-time operations. From an archaeological perspective, SA emerges as both a condition and an effect of the need to make the human part of our socio-technological systems more susceptible to analysis.

6.1 Introduction

Situation awareness (SA) has become a ubiquitous object of knowledge over the last 25 years. SA presents an orientation for academic research, a description of someone's mental state or processes, specifications for interface design, an objective for national (health) policy, a focal point for human factors training, an entry in operating manuals, an explanation for accidents, and nowadays, the lack of it presents a potential reason to be criminally prosecuted and convicted (see, Breakey, Winsen, Dekker, 2015). The broad and diverse use of SA makes it difficult to provide a commonly accepted definition of the object, however, SA generally refers to a person's understanding of a situation and its development. Historically, SA has originated in the air force, as an awareness of the enemy or situation: "SA was recognized as a crucial commodity for crews of military aircraft as far back as World War I" (Endsley, 1995b, p. 32). However, the object of SA "did not receive much attention in the technical and academic literature until the late 1980s, but has become a hot topic ever since" (Stanton et al., 2001).

Objects of knowledge, such as SA, are not 'out there' in the world, waiting for science to discover them. Instead of representing some external reality, French philosopher Michel Foucault argues that discursive practices systematically constitute the objects of which they speak (Foucault, 2002a). As such, the objects of our discourses are historically contingent and arbitrary constructions. With his archaeological analyses, Foucault offers a comprehensive approach to understand how particular constructions of the world emerge and pass as 'truth'. Archaeological analysis thus provides an appropriate theoretical framework to explore the current widespread appeal of SA as a discursive object.

Inspired by Foucault's archaeological approach, this paper offers a study of how SA is discursively constituted in the scientific literature—in what domains the object surfaces and by what disciplines it is invoked. As such, the aim of this paper lies not in providing an epistemological or ontological discussion based on a meta-review of the SA literature, but rather presents the first archaeological analysis of this object: it aims to capture where and how SA emerges and discursively functions in the scientific literature.

6.2 Theoretical framework: Foucault's archaeological approach

6.2.1 Foucault's conception of discourse

Over the last decades, 'discourse' has become a widely used term in the social sciences and humanities. One of the reasons for the popularity of the concept might be an increasing dissatisfaction with positivist approaches to the study of complex social problems (Howarth, 2000; Mills, 1997). Michel Foucault presents such an anti-positivist approach to the study of discourse, assuming that there is no straightforward relation between discourse and reality. Rather than seeing discourses mirror or represent reality, Foucault sees discourses "as practices that systematically form the objects of which they speak" (2002a, p. 54). Discourse is constitutive: it structures the manner in which we perceive and understand reality.

> [W]e must not resolve discourse into a play of pre-existing significations; we must not imagine that the world turns towards us a legible face which we would have only to decipher; the world is not the accomplice of our knowledge; there is no prediscursive providence which disposes the world in our favour. We must conceive discourse as a violence we do to things, or in any case a practice which we impose on them; and it is in this practice that the events of discourse find the principle of regularity (Foucault, 1981, p. 67)

Foucault thus argues that the regularities we see in reality are brought about by the regularities of discourse—which we 'violently' impose on reality. Discourse dictates what can be reasonably thought at a particular time and place, as such, they are historically contingent. The question now becomes, why do we constitute the world in one particular way rather than another? In regards to the object of interest to this paper, why do we speak about SA? In what context is the object used, by whom, and why there? This is what Foucault's archaeological approach aims to describe.

6.2.2 Archaeological analysis and the formation of objects

Archaeology is "an inquiry whose aim is to rediscover on what basis knowledge and theory became possible; within what space of order knowledge was constituted…" (Foucault, 2002b, pp. xxi–xxii). Foucault sees discourses as systems of statements, which emerge under historically and culturally specific 'rules' that determine what can meaningfully be thought and said at a particular time and place. "Archaeology tries to define not the thoughts, representations, images, themes, preoccupations that are concealed or revealed in discourses; but those discourses themselves, those discourses as practices obeying certain rules" (Foucault, 2002a, p. 155). These 'rules of formation', as Foucault (2002a) calls them, are thus a feature of the discourse itself. As opposed to hermeneutics, archaeology thus moves away from seeking underlying essences in discourse and moves towards an approach in which discourse is described in its own terms. "It does not treat discourse as *document*, as a sign of something else", but "is concerned with discourse in its own volume, as a *monument*" (Foucault, 2002a, p. 155); hence the name 'archaeology'.

In studying the 'rules of formation of objects', Foucault (2002a) lays out a number of concepts for studying the manner in which particular discursive objects emerge: (A) The ‘surfaces of emergence’, which ask to describe the fields or settings in which the object of analysis emerges. Given this study’s interest in the object of SA, the analysis will focus on the various applied domains in which this object surfaces. (B) The ‘authorities of delimitation’, which refer to those institutions that delimit, define, and specify the object, as well as establish how it should appropriately be used. As this study is concerned with the scientific constitution of SA, here I will focus on the various (groups of) disciplines that define and transform this object for their purposes. (C) The ‘grids of specification’, which refer to the larger (theoretical) frameworks adopted to develop, define, and specify the object of analysis. As this study focuses chiefly on the emergence of SA, for didactic purposes the grids of specification for SA will not be discussed in this paper.

Foucault explains that we should not see objects as ‘discoveries’ that can readily be taken up by any willing (scientific) discourse. The descriptions of the surfaces of emergence and authorities of delimitation fail to capture the complexity of the relations between them. To understand the space in which possible objects can emerge, we need to understand the overlap and relations—the ‘discursive relations’ (2002a, p. 50)—between these places of its of emergence.

In line with Foucault’s archaeological instructions, in order to understand the emergence of SA, this study draws on the description of the surfaces of emergence, the authorities of delimitation, and the discursive relations between them. In other words, it tries to understand the ‘discursive need’ for this object, and subsequently asks how this object will provide us with a new manner to understand the world: with new leverage, new insights, and new solutions to our problems.

6.3 Method

Describing the ‘surfaces of emergence’ and ‘authorities of delimitation’ of SA in order to understand the formation of this object is not a matter of empirical work, in the positivistic sense of collecting data and categorising it for measurements and accurate descriptions. It requires strategic choices for selecting relevant ‘materialities’ (Foucault, 2002a) in which this discursive object emerges. As such, methodologically, this paper is inspired by the systematic literature review, which is characterised by its explicit research approach—the sources and search strategies for literature are made explicit and the criteria for selection and analysis of the studies are uniformly applied (Petticrew, 2001; Pickering & Byrne, 2013).

6.3.1 Selection and inclusion of SA papers

The process and ritual of scientific publication presents a strong 'technology' (Foucault, 1991a) for fixing a particular understanding of the world as 'truth'. The scientific peer-review process, the critical assessment by colleagues from the same field, is chiefly aimed at analysing and judging what is valid knowledge and what is not. As the academic peer-review process thus presumes a high level of acceptance by the scientific community (Schmoch & Schubert, 2008), this study focuses on the construction of SA in peer-reviewed journal articles, as opposed to books, conference proceedings (which are often not as thoroughly peer-reviewed), training documents, policy guidelines, or operating procedures. Similarly, as the number of citations for a publication is an important indicator of scientific recognition (Allik, 2013; Durieux & Gevenois, 2010; Schmoch & Schubert, 2008; Sharma, 2012), this study focuses on the top cited peer-reviewed articles concerned with the object of SA.

Scopus, the largest citation database of peer-reviewed literature (Aghaei Chadegani et al., 2013; de Moya-Anegón et al., 2007; Elsevier, 2014), was used to select the literature for analysis. The search instructions were on the algorithms "situation awareness OR situational awareness" in the 'Article Title', 'Abstract', or 'Keywords'. At the time, May 2014, this returned 7.390 hits. To include only the peer-reviewed journal articles, these results were further limited by the additional search criteria of 'source type' = 'journal', and 'article type' = 'article' or 'review'. In order to obtain as much diversity as possible in terms of the surfaces of emergence and the authorities of delimitation, no further selection criteria were introduced[14]. This left 1.551 papers that were subsequently

[14] An alternative search strategy, which would have yielded a higher percentage of papers chiefly concerned with the object of SA, was considered. As a more selective search strategy, this would first locate the journals that have published the most papers on the topic of situation awareness—according to Scopus the journal of Human Factors then published 51 papers, Ergonomics 33 papers, Aviation Space & Emergency Medicine 33, IJAP 32, Safety Science 26, etc. on the topic of SA—and consequently select only the top cited SA papers of only these journals. This approach, however, would overlook a vast percentage of highly cited SA research that is published in more domain-specific journals (e.g. the Journal of Anaesthesiology), as opposed to the more generic human factors and psychology journals that typically publish more papers on the topic of SA. For this reason, no pre-selection on the basis of journal was made.

sorted based on the number of 'citations'. Using a minimum of 38 citations as a selection criterion made for a data set of 98 papers, which was considered to be a good set size for this study; analysing more papers was anticipated to yield marginalized returns in providing new insights in terms of surfaces of emergence and authorities of delimitation. These 98 documents provide the materialities for analysis.

6.3.2 Analysis of SA literature

A 'good quality' systematic literature review starts with clear questions or categories for analysis (Petticrew, 2001). In this paper, the categories for analysis are informed by Foucault's suggestions for the archaeological study of the formation of objects. As such, all the selected papers were methodically analysed on the basis of the following archaeological categories: (a) The 'surface of emergence' of SA; (b) The 'authority of delimitation' for SA; and (c) The reasons for invoking SA.

In qualitatively describing and analysing the papers according to these three main categories, I used some heuristics that are based on the archaeological principle to study "discourse in its own volume", not "as a sign of something else" (Foucault, 2002a, p. 155). To categorise a paper's surface of emergence, I looked at the domains of application that the papers are explicitly addressing their research to, specifically in terms of SA. The attribution of a paper to a specific surface of emergence is thus based on the paper's explicit wording of the context in which SA is invoked. This 'bottom up' approach allowed categorisation to be as specific as the papers themselves wished to be.

The analysis of the papers' authorities of delimitation followed a similar strategy. However, because the study is limited to the scientific conception of SA as an object, the analysis of the authorities of delimitation focused on what scientific discipline or larger academic group is speaking. Unfortunately, as journal articles typically do not explicitly state from what scholarly tradition they are representing—that is, they often do not explicitly declare their academic discipline or scientific orientation—the assessment of the authorities of delimitation often asked for more clues about what scientific authority is speaking. Consequently, in addition to the papers' explicit declarations in regards to their academic tradition, categorising the authority of delimitation was informed by the journal in which the paper is published. Another heuristic for assessing a paper's authority of delimitation was presented by the authors' professional affiliations, which are typically provided in journal articles.

In order to limit the complexity of the categorisations, each paper was attributed to only the most evident—most strongly emphasised by the papers themselves—surface of emergence and authority of delimitation. The third category for analysis, which asks about the reasons for the paper to invoke SA, however, was only recorded for the papers that explicitly addressed this question. For this third category, however, not only the most manifest but all the explicit reasons that a paper provides for invoking SA were marked.

6.3.3 Synthesis and presentation of the data

The papers' self-proclaimed descriptions for the surfaces of emergence and authorities of delimitation were collapsed into more general categories, which allowed for some simple quantification: the papers' surfaces of emergence and authorities of delimitation were summarised in pie-charts, as these are a useful tool to present large amounts of data into easily apprehensible illustrations. However, given that this study promotes a qualitative approach, it is of utmost importance to see these pie-charts of 'category counts' for what they are—as merely an illustration—and not to attach any additional value to them. To do so would not only undermine Foucault's intentions for his archaeological projects, it would be 'silly' for multiple reasons.

First of all, the data set is neither exhaustive nor representative. Not only is there no (online) reference database that indexes all the papers that have been written about SA, but even the choice to select papers on the basis of citation is prone to all sorts of distortions: recent papers will be underrepresented, just as the older papers that have been published before digital journals and databases became standard. Also, because the number of citations for these papers is in constant flux, the group of top cited papers is an unstable data set. As such, it is difficult to assess to what extent this way of sampling is representative for the larger cohort of SA papers.

A second issue is that of categorising, which is necessarily subjective. Given Foucault's explicit rejection of positivism, rather than presenting quantitative measures of the validity and reliability of my categories[15], this paper aims to be upfront and transparent about the analytic choices and processes that constitute the analysis. The study's use of categorisation and quantification are only 'tools' to help structure the qualitative description of the various statements about SA. In providing an 'authentic account' (Golden-Biddle & Locke, 1993), the results of my analysis are presented qualitatively, both in terms of process and outcome. Beyond the elaboration on process, the various papers attributed to a particular category will be extensively referenced. This approach leaves critical readers with the possibility to (re)assess my analytic attributions.

The qualitative description of the data will be structured according to the main categories for analysis. After presenting some general results, the paper will describe the various domains in which SA has surfaced, the authorities of delimitation that invoke this object. The discussion of the 'discursive regularities' and the reasons for invoking SA will be based on the numerous quotes that the papers offer in regards to the third category for analysis. Finally, I will present some archaeological reflections about the discursive need and functioning of SA, before concluding with some cautionary remarks about some of the effects that this object may produce.

6.4 General reflection on the results

The broad search and inclusion criteria made for a data set that includes papers from a wide range of domains and disciplines. However, SA is not the main object of interest in all these papers. Some papers are chiefly interested in SA, developing or measuring it for their specific purposes, whereas in other papers SA is merely a sub-topic (if even that), used to introduce the main topic of those papers. As such, it is important to note here that my analysis focused on the parts concerned with SA in these papers. From an archaeological perspective, even the papers that are only marginally concerned with SA play their role in constituting and legitimising it as a discursive object. As the object is used in a particular configuration, with a particular goal and meaning, it is delimitated, specified, and reinforced as an object. Three of the papers originally included in the data set were excluded because it was impossible to establish how SA was to be understood.

[15] Both terms—validity and reliability—are chief indicators of Western science's commitment to a positivist epistemology in which there is only one correct understanding of the world. Such measures are thus far removed from an archaeological approach to the analysis of discourse.

As an example, one paper (Knight, Harnett, & Titov, 2005) accidentally made it into the original data set because a sentence in the introduction ends with 'situation' and the next starts with 'Awareness', thus meeting the search criteria for 'situation awareness'.

Archaeologically it is of no significance whether human factors, health care, or aviation are 'accurately' represented in the data set. Representation is an artefact of various analytic choices and other contingencies (see previous section). Possibly because some domains are not as suitable for research, or because certain disciplines do simply not publish as much research, or they do so in journals with low impact factors, some surfaces of emergence and authorities of delimitation have not made it into the data set for analysis. An evident domain that is missing, for example, is fire fighting. On the other side of the spectrum, the analysis of the papers shows strong 'preferences' for certain domains or disciplines: probably due to the high impact factors of medical journals, healthcare is strongly 'represented'. From an archaeological perspective, this does not mean that health care is the most prominent and important domain in constituting SA; it just means that the object of SA has surfaced there, quite extensively, and is part of the medical discourse (nothing more, nothing less).

6.5 Surfaces of Emergence

In analysing the literature, attributing a paper to a specific surface of emergence was reasonably straightforward, as most papers explicitly locate their problem statement or application of SA in a particular domain. The biggest challenge proved to be combining the plethora of highly specific domains, as worded by the papers themselves, into the larger groups that provide the structure for qualitative description. Choosing whether a study concerned with the SA of unmanned aerial vehicles (UAV) operators should be labelled as primary belonging to 'the military' domain, 'robotics' or even 'aviation' is not always evident. However, given that archaeology is not concerned with quantification and representation, this was deemed unproblematic as long as these kinds of subtleties are qualitatively captured. Keeping these reservations in mind, figure 6.1 displays the grouped surfaces of emergence of the 95 papers included in the study.

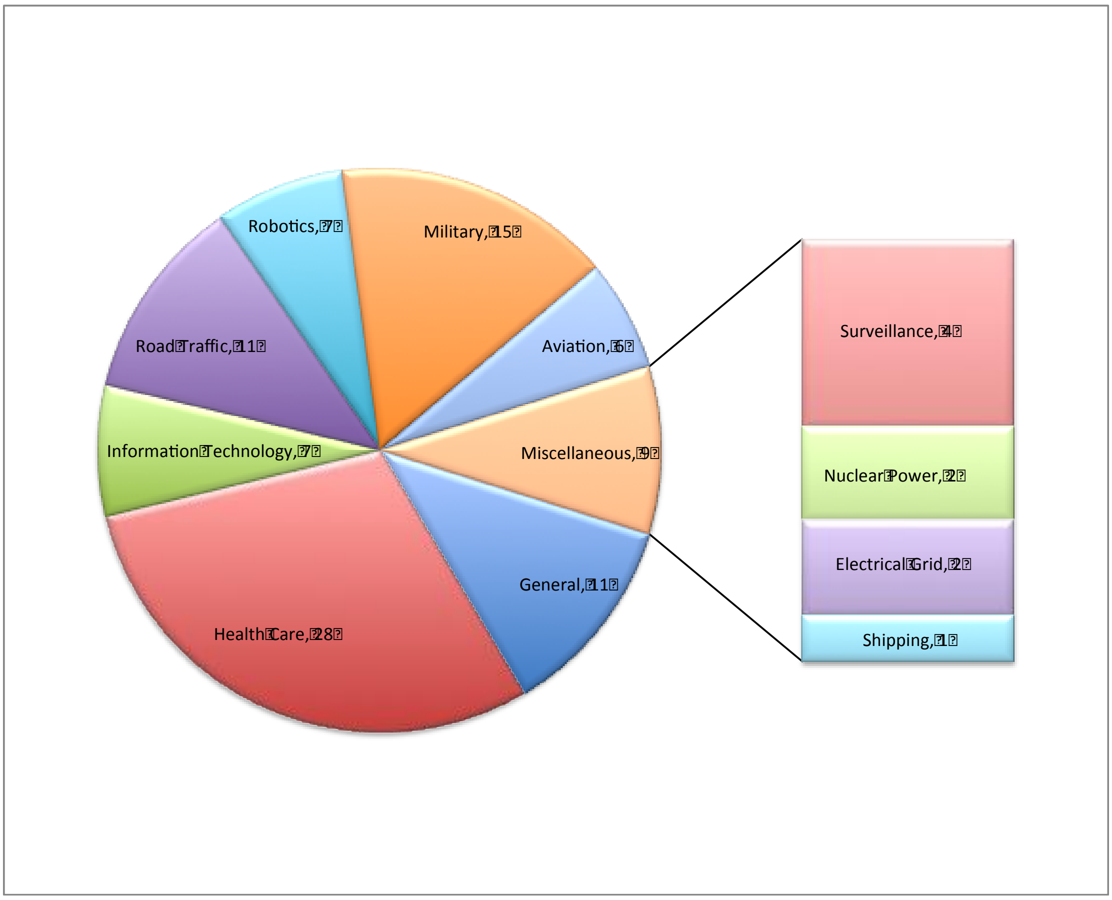

Figure 6.1. The surfaces of emergence of SA.

6.5.1 General

The twelve papers captured under the label 'general' are chiefly concerned with locating SA in a larger, often theoretical, argument that explicitly crosses various domains. In developing her three-level model for SA, Endsley (1995b, p. 37), for example, explains "SA is a broad construct that is applicable across a wide variety of application areas…". As such, the papers in this group (Endsley, 1995a, 1995b; Flach, 1995a; Salas, Prince, Baker, & Shrestha, 1995; Stanton et al., 2001; Wickens, 2008) are deliberately not locating their application of SA in one specific setting or domain and, as such, these papers have not been attributed to a specific surface of emergence for SA.

A second part of this group is comprised of papers that are also explicitly interested in providing a general—not domain specific—argument. However, instead of being chiefly concerned with SA, these papers are all explicitly concerned with the introduction of automation in our socio-technological systems (Endsley & Kaber, 1999; Endsley & Kiris, 1995; Kaber, Onal, & Endsley, 2000; Miller & Parasuraman, 2007; Parasuraman & Wickens, 2008). Finally, one paper that crosses domains invokes SA from a neuropsychological perspective (as an authority of delimitation) to introduce the cognitive process of 'spacial updating' as its central topic (Wolbers, Hegarty, Büchel, & Loomis, 2008).

In light of this study's goal of mapping the emergence of SA as a scientific object, when papers set out to provide a general argument but apply or test their argument in a particular setting, I classified these papers in favour of the more domain-specific groups. Stanton et al. (2006), for example, develop the general notion of 'distributed SA' (DSA), however, as they assess their concept in a military command and control simulation, the paper has been categorised under the military surface of emergence.

6.5.2 Health care

Figure 6.1 shows the large amount of papers that can be attributed to the overarching label of 'Health Care', which is used to depict a variety of medical specialties. In this group, besides generally referring to health care (Campbell, Guappone, Sittig, Dykstra, & Ash, 2009; Catchpole, Mishra, Handa, & McCulloch, 2008; Eva & Regehr, 2007; Gurses & Xiao, 2006; Helmreich & Davies, 1996; Moulton, Regehr, Mylopoulos, & MacRae, 2007; Singh, Petersen, & Thomas, 2006; Weaver et al., 2010; Wright, Taekman, & Endsley, 2004), at least seven papers locate their use for SA in the domain of surgery (Flin et al., 2007; Guerlain et al., 2005; Mishra, Catchpole, Dale, & McCulloch, 2008; Sevdalis, Davis, et al., 2008; Sevdalis, Forrest, Undre, Darzi, & Vincent, 2008; Yule et al., 2008; Yule, Flin, Paterson-Brown, Maran, & Rowley, 2006), three are specifically concerned with SA in anaesthesiology (Fletcher et al., 2003; Gaba, Howard, & Small, 1995; Yee et al., 2005), and then there are papers that place their work in the specific domains of intensive care (Gorman et al., 2000; Reader, Flin, Lauche, & Cuthbertson, 2006), perinatal care (obstetrics) (Miller, Riley, Davis, & Hansen, 2008), emergency care (Beach, Croskerry, & Shapiro, 2003; France et al., 2005), and emergency medical dispatch (Blandford & Wong, 2004). Finally, also captured under the umbrella term of 'health care', are three papers that explicitly locate SA in public health (Bradley, Rolka, Walker, & Loonsk, 2005; Bronstein et al., 2009; Bronstein, Spyker, Cantilena Jr, Rumack, & Dart, 2012).

Often these health care papers present SA as a 'non-technical skill' that medical staff needs to be trained and assessed for, mostly in light of 'patient safety' concerns, either in real life or simulated health care scenarios. "During the last 20 yr, the importance of non-technical skills for delivering safe and high-quality medical care has been increasingly recognized, if not explicitly addressed, in medical training" (Reader et al., 2006, p. 551). SA is said to be a relevant non-technical skill across the various medical domains (e.g. Gaba et al., 1995). However, in acknowledging that "each working environment has its own unique non-technical skill requirements" (Reader et al., 2006, p. 551), some papers argue that SA is of particular importance in certain medical settings: "The operating room (OR) has a unique set of [conditions], this complex environment provides multiple opportunities for unclear communication, clashing motivations, and errors arising not from technical incompetence but from poor interpersonal skills" (Catchpole et al., 2008, p. 699).

6.5.3 Military

The military domain is a second surface of emergence that is well represented in the data set. "Because of the complexity and dynamic nature of the command and control environment, maintenance of situational awareness is considered to be of utmost importance" (Cummings, 2004, p. 656). Many paper use SA in 'command and control' settings (Artman, 2000; Cummings, 2004; Gorman, Cooke, & Winner, 2006; Kaempf, Klein, Thordsen, & Wolf, 1996; Sonnenwald & Pierce, 2000; Stanton et al., 2006), also referred to as 'C2' or even 'C4i'—"command, control, communication, computers and intelligence (C4i)" (Salmon et al., 2009, p. 225).

A few other military enterprises are grouped under this label. Besides the development of military equipment—two papers work on the fusion of visible and thermal imagery to improve SA (Toet & Franken, 2003; Toet, Ijspeert, Waxman, & Aguilar, 1997)—the development and control of UAVs or other ROVs (remotely operated vehicles) seems especially concerned with SA: "many human interface challenges have arisen in the ROV domain, potentially contributing to reduced situational awareness (SA) and degraded mission performance" (Ruff, Narayanan, & Draper, 2002, p. 336). Driven largely by military pursuits and funding, the application of these technologies is mostly in military settings. As such, the papers concerned with the development of such technologies (Parasuraman, Cosenzo, & De Visser, 2009; Ruff et al., 2002) were assigned to the military surface of emergence, even though the extensive description of their autonomous functioning could also qualify them for the domain of robotics. Alternatively, their aerial orientation could have qualified them for the 'aviation' group. For similar reasons, the papers that invoke SA in talking about 'combat aircraft' (Svensson, Angelborg-Thanderz, Sjöberg, & Olsson, 1997) or the use of 'HUDs' (head up displays) for reducing spatial disorientation (Rupert, 2000), were allocated to the military (aviation) domain, rather than (commercial) aviation. Finally, two papers explicitly locate their simulated combat studies in the domains of the air force (Jones et al., 1999) and the navy (Randel, Pugh, & Reed, 1996).

6.5.4 Aviation

A vast number of papers explain that aviation has always been at the forefront of developing SA: "Historically, the aviation community is responsible for introducing the research community to a pervasive phenomenon of tactical flight operations called situation awareness (SA)" (Gorman et al., 2006, p. 1314). However, as illustrated by 6.1, the number of papers that are primarily concerned with SA in aviation is surprisingly small compared to healthcare and military settings. The six papers that locate their argument predominantly in aviation refer to aviation in general (Jones & Endsley, 1996; Wickens, 2002), commercial aviation (Adams et al., 1995; Smith & Hancock, 1995), or air traffic control (ATC) (Kaber, Perry, Segall, McClernon, & Prinzel Iii, 2006; Remington, Johnston, Ruthruff, Gold, & Romera, 2000).

The small number of papers attributed to the aviation group may be the effect of sampling issues; aviation is a domain that may predominantly publish in papers with lower citation indexes. Alternatively, the exclusion of the papers that were primarily attributed to 'military aviation' or the 'general' group may also be a factor. Despite the low number of papers primarily attributed to the aviation group, the importance of SA in aviation—"SA has grown in importance as a major design goal for civil, commercial, and military aircraft" (Endsley, 1995b, pp. 32-33) —and the leading role that aviation has (had) in developing objects as SA—"the coming of age of the research into situational awareness is closely coupled with the increase in the degree of automation in flight control" (Stanton et al., 2001, pp. 189-190)—is acknowledged by papers across the various surfaces of emergence.

6.5.5 Road traffic

With eleven papers, 'road traffic' is a domain that is strongly represented in the data set. This group can roughly be separated into two strands, one that focuses on driver SA in regards to traffic safety (Drews, Pasupathi, & Strayer, 2008; Gugerty, 1997; Kass, Cole, & Stanny, 2007; Ma & Kaber, 2005; Underwood, 2007), and one that focuses on equipment design for road traffic (Petrovskaya & Thrun, 2009; Sengupta et al., 2007; Stanton & Young, 2005; Wang et al., 2011; Won & Langari, 2005). As part of the first strand, a number of studies concerned with SA in traffic safety are concerned with the distracting effects of cell phone use whilst driving. As most of these papers aim to assess whether the use of a cell phone reduces SA, in this group I have also included a study concerned with the distracting effects of pedestrians using cell phones (Nasar, Hecht, & Wener, 2008).

The second strand focuses on the study of new technologies for cars and driving. As these papers talk about highly automated technologies—such as automatic collision warning systems, adaptive cruise control, and even autonomous self-driving cars—some of these papers could have been captured under the banner of 'robotics'. However, given their explicit focus on automobiles, they were attributed to this group. Moreover, as these studies are interested in developing new driving technologies, they are all doing this against a backdrop of safety (as this is an important design aspect of vehicle design). Both strands are thus highly interconnected and share a similar interest in driver (or pedestrian) SA.

6.5.6 Robotics

Nowadays humans use robotic systems in a broad variety of settings, therefore, the group of 'robotics' could be said to present a meta-domain as a surface of emergence. However, even though "human interactions with robots span almost every robotic endeavour, HRI [human robotic interaction] research has been conducted primarily in six application domains: entertainment, museum docents, personal assistants, health care, space exploration, police SWAT teams, military robotics, and rescue robotics" (Murphy, 2004, p. 139). Interestingly, five of the seven papers that constitute the 'robotics' surface of emergence in this study are explicitly concerned with rescue- or 'urban search and rescue' (USAR) robotics (Burke, Murphy, Coover, & Riddle, 2004;

Hsieh et al., 2007; Hsieh, Cowley, Kumar, & Taylor, 2008; Kumar, Rus, & Singh, 2004; Murphy, 2004). Apparently, “the relation between humans and robots in USAR is different than the relation in manufacturing, office, or even security applications of robots” (Burke et al., 2004, p. 87).

This domain specificity in regards to SA may be explained by another aspect that the papers in this group have in common: they all seem to invoke SA because of human-robot interaction problems, mostly in the process of teleoperation: “Even for autonomous and semiautonomous systems, teleoperation will be an important default mode. However, teleoperation can be a challenging task because the operator is remotely located. As a result, the operator’s situation awareness of the remote environment can be compromised …” (Chen, Haas, & Barnes, 2007, p. 1231). Teleoperation thus seems to present a significant challenge to the SA of USAR teleoperators.

Finally, besides the USAR papers and Chen et al.’s (2007) paper that explicitly focuses on ‘human performance issues for teleoperated robots’, the last paper in this ‘robotics’ group (Choi, Brunet, & How, 2009) discusses the coordination of a fleet of autonomous robots, which is a theme that also surfaces in SA papers in other domains (most prominently in regards to ROVs).

6.5.7 Information technology

The surface of emergence labelled as ‘information technology’ covers a range of applications of SA. However, the papers in this group are all concerned with human-computer interactions and the digital management and sharing of information. As such, two papers explicitly locate their research in the field of ‘Computer Supported Cooperative Work’ (CSCW). They both explain that in our modern socio-technological systems “the workforce becomes increasingly distributed across space and time and increasingly mobile, the need to collaborate with remote partners to accomplish collaborative tasks has increased substantially” (Kraut, Fussell, & Siegel, 2003, p. 14). Consequently, SA “has recently begun to receive considerable attention in CSCW and groupware research” (Gutwin & Greenberg, 2002, p. 411).

Less specific then CSCW, the other papers in this group can be linked to software engineering. One paper investigates our modern networked societies and explores the specific field of 'citizen sensing', "that is, humans as citizens on the ubiquitous Web, acting as sensors and sharing their observations and views using mobile devices" (Sheth, 2009, p. 87). Kokar, Matheus, and Baclawski (2009) talk to the 'information fusion community' and aim to provide an 'ontology-based situation awareness'. The other studies do not present a narrow label for their field of application, but like the other papers in this group are generally located in the field of 'human-computer interaction' (Lindeman, Yanagida, Noma, & Hosaka, 2006), 'human-computer interface' design (Selker & Burleson, 2000), and explain "how the interaction with humans is one of the critical components of modern Distributed Artificial Intelligence (DAI)" (Tweedale et al., 2007, p. 1089).

6.5.8 Miscellaneous: surveillance, nuclear power, electrical grid, and shipping

Finally, there is a miscellaneous 'group' of domains that makes up the side bar in figure 6.1. 'Surveillance', as a surface of emergence, is represented by four papers in the data set (Collins, Lipton, Fujiyoshi, & Kanade, 2001; Hampapur et al., 2005; Morris & Trivedi, 2008a, 2008b). Surveillance studies seem to have a keen interest in the object of SA—or 'scene awareness'—as "situational awareness research is to understand the interactions and behaviors present in a scene. This scene awareness is particularly important for visual surveillance systems that must continually monitor a site" (Morris & Trivedi, 2008a, p. 425).

Both represented by two papers in the top cited literature on SA are the industries of nuclear power (Hogg, Folleso, Strand-Volden, & Torralba, 1995; Sebok, 2000) and the electrical grid (Hauser, Bakken, & Bose, 2005; Zhang et al., 2010). The importance of SA in these surfaces of emergence is tied to supervisory control mode of operations that is characteristic of both domains: "The operator of a nuclear power plant must have knowledge of the current process state at all times, and the ability to use this knowledge effectively in predicting future process states and controlling the process to attain operational goals" (Hogg et al., 1995, p. 2394). Finally, shipping, traditionally one of the most important domains for traffic studies is represented by one paper (Hetherington, Flin, & Mearns, 2006) in the top-cited papers on SA.

6.6 Authorities of Delimitation

In the various surfaces of emergence, SA has been delimitated, defined, and further legitimised by a number of different authorities of delimitation. As this study is limited to the scientific conception of SA, the analysis focuses on the particular scientific authorities that delimitate the object. Unlike with the categorisation of the surfaces of emergence, journal papers often do not explicitly state from what academic discipline they are speaking. Information about the journal in which a paper is published, as well the authors' affiliations, were consulted to establish a paper's authority of delimitation. However, the categorisation of papers into discipline-specific categories proved difficult, if not impossible. As an example: an article concerned with driver SA in road traffic, which is published in a generic safety journal, can be attributed to the academic disciplines of 'transportations studies', 'human factors', 'safety science', and, even 'psychology' may be a justifiable allocation depending on the paper's methodological approach or the authors' affiliations. Consequently, categorisation at this level of specificity produced highly volatile categories that were not stable enough to structure the qualitative description.

Instead, the analysis and synthesis of the authorities of delimitation was based on clusters that transcend specific academic disciplines. These clusters present coherent groups of disciplines that have enough stability and distinctive value (i.e. unique characteristics) to allow for qualitative description. Besides outlining the typical disciplines that constitute these clusters, as evidenced by the author affiliations and the journals in which the research is published, the clusters present a typical research orientation and methodological preferences. The distribution of the papers over the clusters—which thus make up the authorities of delimitation in this study—is presented in figure 6.2.

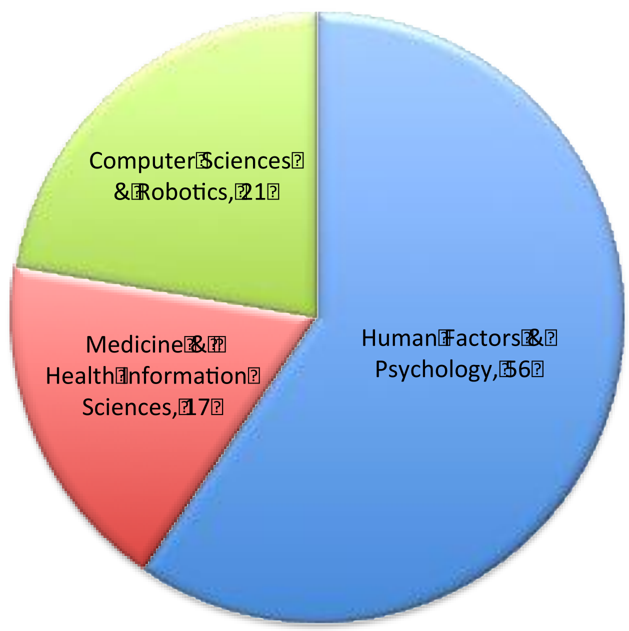

Figure 6.2. The authorities of delimitation for SA.

6.6.1 Human factors & psychology

As figure 6.2 displays, the authority of delimitation attributed to more than half of all the papers is the cluster of 'human factors and psychology'. This cluster comprises most scientific disciplines that take the human subject as their main object of analysis. Psychology—traditionally the academic discipline concerned with the scientific study of (individual) human behaviour and cognition—makes for an evident candidate to have something to say about SA. However, only few papers in the data set explicitly state to be talking from a psychological viewpoint. Similarly, only a few studies are published in typical psychological journals: 'Military Psychology', 'Journal of Experimental Psychology', 'Current Directions in Psychological Science', 'Cyberpsychology and Behavior', and 'Nature Neuroscience'. However, looking at the affiliations of the authors of the papers included in the analysis, a great number of papers are affiliated to 'schools of psychology', 'psychology departments', and other kind of psychological research institutes.

The great number of papers that explicitly label their work as 'human factors' research, which is often conducted from within psychology departments, can explain this discrepancy. The discipline of human factors and ergonomics (hereafter referred to as human factors)—the scientific discipline concerned with the study of the interaction of humans and technology—has historically developed out of the experimental psychology, and can still be characterised by experimental psychology's empirical-experimental methodology. Early human factors pioneers, such as Fitts and Jones (1947), were experimental psychologists who adapted their laboratory techniques to applied real-world problems, and in doing so created a new way of analysing accidents and human performance issues. In addition to the explicit references to a human factors approach, human factors is strongly represented by the vast numbers of papers that are published in typical human factors journals: 'Human Factors' (13 papers), 'Ergonomics' (9 papers), 'Applied Ergonomics', 'International Journal of Industrial Ergonomics', 'International Journal of Human Computer Studies', 'Human-Computer Interaction', and 'Information Processing and Management'.

Some papers hint at the sociological origin of human factors, because "[r]esearch in human factors includes the discovery and application of knowledge about individual *and team* interactions with technology" (Helmreich & Davies, 1996, p. 277, italics are mine). Interestingly, however, none of the authors explicitly claim to be speaking about SA from a sociological tradition, are affiliated to sociology departments, or publish their work in typical sociological journals. Other disciplines concerned with more specific aspects of the interaction of humans and technology, for example in regards to safety (e.g. safety science) or domain specific variants (e.g. transportation studies), were also included in the 'human factors and psychology' cluster. These disciplines, which just like human factors are often located in more traditional departments as that of psychology, are represented in the data set by publications in outlets such as 'Safety Science', the 'Journal of Safety Research', 'Accident Analysis and Prevention', 'Aviation Space and Environmental Medicine', 'Journal of Intelligent Transportation Systems', and 'Transportation Research'. Not only do these disciplines share human factors' interest in the study of human-machine interactions, they often also rely on the empirical-experimental approach, especially when studying objects as SA.

Despite the difficulties in methodically distinguishing between the various scientific disciplines that make up this large cluster of human factors and psychology, it is evident that the various disciplines concerned with studying the interaction of humans and technology is an important authority of delimitation for the object of SA: "The concept of situation awareness lies at the heart of the intersection between basic cognitive psychology and the applied science of human factors" (Wickens, 2008, p. 402).

6.6.2 Medicine & the health information sciences

In addition to the large, but diverse, cluster of human factors and psychology, figure 6.2 displays two other, narrower, authorities of delimitation. Given the strong representation of the health care domain in the surfaces of emergence, it is not surprising that 'medicine and the health information sciences' presents its own cluster here[16]. Even though medicine can also be seen as belonging to the human factors and psychology cluster, there are a number of characteristics specific to this cluster that make it into a distinct authority of delimitation.

The first part of this medical cluster is characterised not only by its focus on the health care domain, but also by the domain specific affiliations of the authors and the outlets for their research. Basically all the papers that belong to this group are written by medical practitioners—mostly affiliated to various hospital departments—for other medical professionals. As such, all these papers are published in medical journals: 'Anesthesiology', 'Annals of Surgery', 'American Journal of Surgery', 'International Journal of Medical Informatics', 'Surgical Endoscopy and Other Interventional Techniques', 'Journal of the American College of Surgeons', 'The American Journal of Surgery', 'Quality and Safety in Health Care', 'Academic Emergency Medicine', 'World Journal of Surgery', and the 'Journal of the American Medical Informatics Association'. As a result of speaking chiefly to a health care audience, the medical disciplines differ from those disciplines that make up the human factors cluster, which are characterised by

[16] Even though the medical disciplines are chiefly concerned with health care, this does not mean that they are the only authority that has something to say about this domain. To see this, consider that 28 papers were labelled as belonging to the domain of health care whereas 'only' 17 papers were attributed to the medical disciplines or health information sciences. The cluster of human factors and psychology presents another prominent authority of delimitation for health care. For human factors, health care is just another domain struggling with the typical human factors issues of people interacting with technology. Such human factors studies get published in typical human factors journals as 'Human Factors' and the 'International Journal of Human Computer Studies' as well as in health care outlets such as 'Medical Education', 'Surgeon', 'World Journal of Surgery', and the 'British Journal of Anaesthesia'.

more generic and multi-disciplinary sciences with a broader (i.e. more heterogeneous) audience. Compared to human factors' rigorous and elaborate empirical-experimental method, medicine's homogeneous audience allows the medical authorities to be more straightforward and pragmatic in their theoretical and methodological specification of SA.

A second part of the medical cluster is that of the 'health information sciences'. Contrary to the other medical disciplines, researchers in the health information sciences are typically not medical practitioners, but are affiliated to research institutions as 'Schools of public health', (academic) health centres, national patient safety agencies, and institutes as the 'American Association of Poison Control Centers'. As an authority of delimitation, this group is characterised by its research method: the statistical analysis of various societal demographics or (large) medical data sets. Through the statistical tracking of public health events—(food) poisoning, disease outbreaks, and even possible biologic terrorism attacks—these disciplines are concerned with 'population surveillance'. Similarly, in statistically analysing large medical data sets, the health information sciences intend to provide guidance on medical procedures, tools, and interventions, most prominently in regards to patient safety concerns.
The papers belonging to the health information sciences publish their work in outlets ranging from more general medical journals to discipline specific journals[17]: 'Quality and Safety in Health Care', 'Clinical Toxicology' (2 papers), and 'MMWR' (Morbidity and mortality weekly report), 'Academic Medicine', 'Journal of General Internal Medicine', 'Journal of Perinatal and Neonatal Nursing', 'Surgeon', and 'Surgical Endoscopy and Other Interventional Techniques'.

6.6.3 Computer sciences and robotics

Given their shared focus on humans and human interaction (either with technology or with other humans), the medical disciplines and health information sciences can be seen as a sub-group of the large human factors cluster. Conversely, the 'Robotics and Computer Sciences' cluster provides a more distinctive authority of delimitation as it

[17] An interesting exception here is a paper (Gorman et al., 2000) published in 'Library Trends', whose authors are affiliated to 'the health sciences university' and study a topic that is characteristic for this cluster—electronic bundles for physicians—however, they do so through a for the health sciences atypical methodological approach of field observations.

takes technology as its central subject. Illustrated by the author affiliations—departments of computer sciences or electrical engineering—this authority of delimitation is constituted by a vast range of young disciplines and research orientations, such as information (technology) sciences, software development, artificial intelligence, and robotics research. Just as it is both difficult and archaeologically irrelevant to assess where human factors ends and psychology begins, I did not even consider separating the computer sciences from disciplines as robotics or electrical engineering.

It was often easy to ascribe papers to this cluster, as in addition to the straightforward author affiliations, most of the journals in which the various disciplines of robotics and the computer sciences publish are easily identifiable as belonging to the computer sciences: In addition to eleven publications in various journals of the IEEE (the Institute of Electrical and Electronics Engineers), the papers that make up this cluster were all published in typical robotics and computer science journals: the 'Journal of Field Robotics' (2 papers), 'Autonomous Robots', 'Journal of Intelligent Transportation Systems: Technology, Planning, and Operations', 'IBM Systems Journal', 'Virtual Reality', 'Journal of Network and Computer Applications', 'Computer Supported Cognitive Work', 'Journal of Network and Computer Applications', 'AI Magazine', 'Information Fusion', 'International Journal of Computational Intelligence Systems', and the journal of 'Human-Computer Interaction'.

Papers that were categorised as belonging to the 'robotics' surface of emergence are not necessarily delimitated by the 'robotics and computer sciences' authority of delimitation. For example, two robotics studies that are predominantly concerned with the interaction of the human operator and the robotic system—as evidenced by their titles, 'Moonlight in Miami: A Field Study of Human-Robot Interaction' (Burke et al., 2004) and 'Human performance issues and user interface design for teleoperated robots' (Chen et al., 2007), and the respective journals in which these papers are published: 'Human-Computer Interaction' and 'IEEE Transactions on Systems, Man and Cybernetics'—were attributed to the human factors cluster rather than that of the computer sciences.

Interestingly, the two strands of research application that make up the surface of emergence labelled as 'road traffic'—traffic safety and equipment design for automobiles—are delimitated by respectively the human factors and psychology and robotics and computer sciences clusters. Conversely, all the papers from the military domain, even the ones concerned with equipment design, are authorised by the human factors cluster. The same holds for the 'aviation', 'shipping', and 'nuclear power' domains. Finally, the 'surveillance' and 'electrical grid' surfaces of emergence are entirely authorised by the computer sciences.

6.7 Common characteristics: the discursive relations

After having outlined the surfaces of emergence in which SA has appeared, and the authorities of delimitation that specify this object, it is apt to ask what features and goals they share. To understand the space in which SA emerge, we need to understand the 'discursive relations' between them. In this section I argue that the surfaces of emergence of SA, as well as the concerns of the various authorities of delimitation, are—often explicitly—characterised by high levels of complexity, systems control tasks, and a central concern for risk and safety.

First, systems complexity is typically regarded as a function of "the number of system components, the degree of interaction between these components, and the dynamics or rate of change of the components" (Endsley, 1995b, p. 53). To show the widespread concern with complexity that the papers in the data set share, I provide just one example from each surface of emergence in table 6.1.

Table 6.1

Concern with complexity per 'surface of emergence'

Health care	"The operating room (OR) has a unique set of team dynamics … this complex environment provides multiple opportunities for unclear communication, clashing motivations, and errors arising not from technical incompetence but from poor interpersonal skills." (Catchpole et al., 2008, p. 699)
Military	"C4i environments pose a great challenge to the HF community. As described previously, the environment is typically complex, dynamic and information rich." (Salmon, Stanton, Walker, & Green, 2006, p. 228)
Aviation	"Just as the aircraft and its airspace are complex and dynamic, so also are the

	systems within the aircraft.” (Wickens, 2002, p. 131)
Road traffic	“Driving behavior has been becoming increasingly complex in urban traffic.” (Wang et al., 2011, p. 302)
Robotics	“The rescue enterprise involves a diverse set of people, from ordinary citizens that are victims of a disaster to highly trained rescue professionals on through robot specialists. … The complexity of relationships and tasks in a rescue enterprise pose challenges…” (Murphy, 2004, p. 138)
Information technology	“From a software engineering perspective, one would expect to gain major benefits from intelligent agent technology through its deployment in complex distributed applications such as virtual enterprise management and the management of sensor networks.” (Tweedale et al., 2007, p. 1100)
Surveillance	“Ensuring high levels of security at public access facilities like airports and seaports is an extremely complex challenge.” (Hampapur et al., 2005, p. 83)
Nuclear power	“The operators of large, complex systems such as flexible manufacturing systems, refineries, and nuclear power plants must also rely on up-to-date knowledge of situation parameters to manage effectively.” (Endsley, 1995b, p. 33)
Electrical grid	“The electric power infrastructures in North America and Western Europe are among the most complex systems ever constructed.” (Hauser et al., 2005, p. 47)
Shipping	“We chose the electronic warfare technician (EW) operator as our area of study because the complexity of the task , time constraints for decision making, and the variety of information sources which must be synthesized rapidly … EW operators working on ships use high tech computer equipment to help detect and identify the signals from numerous potentially hostile radar systems.” (Randel et al., 1996, p. 581)

In addition to complexity, a second shared characteristic of the various planes of emergence seems to be the centrality of ‘realtime systems control’ tasks. Endsley and Kaber (1999, p. 464) studied “a wide array of cognitive and psychomotor tasks requiring realtime control within numerous domains including air traffic control, aircraft piloting, advanced manufacturing and teleoperations.” Besides corroborating the argument that these domains are characterised by complexity, of interest here are the functions that Endsley and Kaber see as intrinsic to all these domains: “(1) *monitoring*—scanning displays to perceive system status; (2) *generating*—formulating options or strategies for achieving goals; (3) *selecting*—deciding on a particular option or strategy; and (4) *implementing*—carrying out the chosen option”. (p. 464)

These characteristics for 'systems control' tasks map directly onto the surfaces of emergence for SA. In all the domains concerned with SA, operators are specifically tasked with 'monitoring' various displays to perceive system status, whether the system's parameters are a patient's vitals, the temperature of a nuclear reactor core, aircraft moving through controlled airspace, or the surveillance feeds of CCTV security cameras. Based on the perceived system status, operators need to generate, select, and implement various strategies to support or maintain operations. SA is of interest to any domain concerned with systems control tasks, in which operators are required to stay 'up-to-date' with a complex and dynamically changing environment (Gaba et al., 1995): "In essence, any task that requires people to keep track of events is a potential candidate for situational assessment research and application" (Stanton et al., 2001, p. 193).

A third relation between the various surfaces of emergence and authorities of delimitation, which is directly connected to the characteristics of complexity and systems control, is a central concern with safety. Beyond the traditional 'high-risk' industries as aviation, nuclear power, health care, military operations, and other transportation systems, even the papers studying the electrical power infrastructure are doing so because they are "increasingly operating nearer to their safety limits" (Hauser et al., 2005, p. 47). Similarly, due to the key role that information technologies (including robotics) play in our modern societies, these modern domains also have a prime concern with reliable and safe operation. Server outages, for example, may have disastrous (not only in financial terms) and widespread effects in our networked societies. As such, the outlined surfaces of emergence for SA could convincingly be labelled as 'safety-critical' industries or systems: "In essence, SA is … a critical concept in any field that involves complex and dynamic systems where safety is a priority" (Singh et al., 2006, p. 159).

6.8 Reasons for invoking SA

Now that we have seen where SA emerges, the question becomes why SA emerges here. That is, why does SA surface in these domains, characterised by complexity, systems control, and a concern for safety? Why are the outlined authorities of delimitation concerned with these characteristics? The discussion presented in this section is supported by the various reasons that the papers present for invoking SA (see section

3.2). This section presents some of the recurring 'themes' for invoking SA, through which this section aims to provide a backdrop to understanding the discursive need that SA aims to fulfil.

6.8.1 The introduction of automation

Complexity is not only one of the shared characteristics of the fields in which SA has surfaced, it is also one of the central 'problems' that the SA papers are trying to address. A number of papers attribute this complexity to technical developments, most prominently to the widespread introduction of automation in our modern socio-technological systems—"that is, using a device, machine, or system that accomplishes, fully or partially, a function that was or could be performed by a human" (Miller & Parasuraman, 2007, p. 57). These papers explain that with the start of the computer age, technology in the form of automation started to amplify increasingly complex cognitive tasks. "Automation is being increasingly applied to higher-level decision tasks in the form of artificial intelligence (AI) or expert systems" (Endsley & Kiris, 1995, p. 385). As such, the pervasive introduction and effects of these 'advanced automation technologies' is of central concern to most of the papers in the data set.

Historically, the introduction of automation was largely driven by the desire to reduce costs through the reduction of human workload and consequent staffing requirements, "such efforts usually assign a computer or mechanical controller to perform those tasks technically possible, and remove human operators from the control loop by placing them in the job of system monitor" (Endsley & Kaber, 1999, p. 463). As the tasks automated in system design were those that could most easily be automated (Sebok, 2000), automation was largely driven by technical possibilities: "designers are apparently tempted by the possibilities, created by modern computer technology, to include increasingly complex and numerous options (modes) and displays in their systems" (Svensson et al., 1997, pp. 363-364). As a result of this 'technological imperative', many papers point to the finding that automation does not take into consideration the needs and limitations of its human end-users (Endsley & Kaber, 1999; Kaber et al., 2000; Tweedale et al., 2007). This problem is also been referred to as 'clumsy automation' (Sebok, 2000).

Given the advanced state of today's technology, numerous human activities can, in principle, be automated, however, the "question is no longer whether one or another function can be automated but, rather, whether it should be" (Wiener and Curry, 1980, as read in, Parasuraman & Wickens, 2008, p. 517). Across the various SA papers, this question is often posed in terms of LOAs—'Levels of Automation', specifying the degree to which a task is automated. A high LOA is represented by the case that automation works (almost) autonomously, and people merely have to monitor (the output of) the system. However, as many papers acknowledge, even though fully automated functioning may be technically possible, it may not always be desirable.

Numerous papers explain that even though automation is employed to simplify operators' tasks so as to reduce workload, it rather "changes" the nature of work in our socio-technological systems (Ma & Kaber, 2005, p. 940; Ruff et al., 2002, p. 336). The changes that come with the introduction of automation present a new range of questions and consequences for our modern socio-technological systems. Besides the ethical issues—"the passive role of monitoring an automatic system is less satisfactory from a human performance perspective than the active role of controlling it" (Stanton & Young, 2005, p. 1311)—the papers in the data set raise a number of practical problems that come with the ubiquitous introduction of automation in our socio-technological systems. Presented as effects of automation, the problems of 'information overload', 'workload ironies', and the 'distancing of operators' from their work are recurring issues.

6.8.2 Effects of Automation: Information overload

As modern technology makes it feasible to provide operators with a wealth of information, one factor that is often reported as both the cause and the effect of the increased complexity of our automated systems is the vexing amount of information that these systems can produce (Gorman et al., 2000; Gutwin & Greenberg, 2002; Stanton et al., 2006). Adams et al., for example, explain that, "pilots may be simultaneously responsible for the information presented by up to 400 separate gauges and instruments. They are, moreover, responsible for the interpretation of that information" (Adams et al., 1995, pp. 85-86). In addition to the information complexity that our current systems already produce, many papers point out that the information loads of our systems will only keep increasing: "computing resources and network bandwidth continue to increase … , users will have to deal with an increasing amount of information" (Lindeman et al., 2006).

In terms of SA, most papers are not as concerned with the increasing amounts of information *per se*, as with people's ability to process and act upon these perplexing amounts of data. "Large amounts of data are generated, making it infeasible for a human to accurately process" (Morris & Trivedi, 2008a, p. 425). In their paper, appropriately entitled 'Information Complexity', Svensson et al. explain that people working in our modern automated "systems have to process a considerable amount of complex information, much more comprehensive than in older systems. The information and decision making processes have become more and more demanding and the risk of mental overload has increased" (1997, p. 364). Similarly, Sonnenwald and Pierce also point to the effect that the increasing complexity of information has on people's ability to work effectively: "Increasingly, as the diversity and complexity of the battlefield and complexity of information pertaining to the battlefield increases, no single individual can acquire and process the diverse and often rapidly expanding information needed to create and execute battle plans effectively" (Sonnenwald & Pierce, 2000, p. 462)

This particular problem of information *overload* is a returning issue in the SA papers. "In complex and dynamic environments, attention demands resulting from information overload, complex decision making, and multiple tasks can quickly exceed a person's limited attention capacity" (Endsley, 1995b). Burke presents an example that is typical for all the surfaces of emergence of SA, characterised by real-time systems control tasks: "searching, mapping, interpreting what is being seen on the video monitor, making decisions about what to do with that information, and dealing with the physical stresses of the job may be overloading the operator in the USAR environment as well" (Burke et al., 2004, p. 109).

6.8.3 Effects of automation: workload ironies

The information overload problems, outlined above, are often related to operator workload problems: "the workload of the … operator can become extremely high because of the large amount and fragmented nature of the information inputs, and because of the varying reliability of some of these sources of information" (Randel et al., 1996, p. 581). In regards to SA, information overload seems to become problematic in combination with time pressure since “operators need to integrate the pieces of information about the situation they receive from different sources, over a period of time, and through different modalities. Furthermore, when presented with major incidents, complications include ambiguous information and significant time pressures” (Blandford & Wong, 2004, p. 448).

In addition to the increase in workload as an effect of the increase in the information load, a frequent concern is the observation that (clumsy) automation itself could actually increase workload, particularly during non-normal operation: “Many current automation approaches have been shown to actually degrade human and/or system performance … , including increasing operator workload” (Ruff et al., 2002, p. 336). As automation is typically employed to reduce operator workload, many papers refer, either explicitly or implicitly, to Bainbridge’s ‘ironies of automation’ (see Bainbridge, 1983), to discuss the ‘performance costs’ of automation. “That high LOAs could leave an operator bored is, perhaps, understandable, but that automation can increase workload is one of the ‘ironies of automation’” (Miller & Parasuraman, 2007, p. 59).

In other words, automation may decrease operators’ physical workload at the cost of increasing their cognitive workload, especially during periods in which cognitive workload is already high. Highly automated cockpits, for example, may have reduced the amount of manual work for flight crews, however, the ‘mental discipline’ to oversee these automated functions may have been increased (Adams et al., 1995). As another ‘irony of automation, “when workload is the highest, automation is often of the least assistance, as it usually can handle only routine tasks” (Endsley & Kiris, 1995, p. 384).

6.8.4 Effects of automation: distancing the operator and work

In addition to problems with workload and information overload, some papers explain that the introduction of automation changed the nature of the demands and responsibilities of the operator: “automation does more than simply replace human activity. It actually changes operator functional requirements as well and can potentially impose new coordination and monitoring demands” (Ruff et al., 2002, p. 336). Increasingly, the role of the operator is that of system monitor—which, we have seen, is characteristic of the domains in which SA has surfaced. The rearrangement of functions between the operator and the automation has some problematic consequences: “This traditional allocation of functions to human and computer has been associated with problems in operator effectiveness in overseeing automated system functioning and intervening in system operations by taking control from a computer during failure modes" (Kaber et al., 2000, p. 410). In addition to pointing out that the properties of human cognition are generally ill-suited for vigilance tasks, the papers in the data set raise some recurring concerns about the operator's role as monitor.

A prominent issue attributed to the rearrangement of functions between the operator and the automation is the detachment of the operator from real-time operations. By changing the role of operators from being actively involved with operational processes to having them passively monitor the automation, the operator is—often both physically and cognitively—'distanced' from the operational processes that they are responsible for. Computers now mediate the interaction between humans and the current state of the system. Even though automation can display a vast amount of parameters, the distance between the operator and the process that is controlled may eliminate certain critical parameters for optimal performance. "The removal of operators from physical contact with the process may impoverish their diagnostic environment despite the more sophisticated displays" (Moray, 1986, as read in, Endsley & Kiris, 1995, p. 384). In other words, the detachment of the operator from real-time operations has made the system more 'unpredictable', which "refers to the inability of the human to know exactly what the automation will do when. Unpredictability is a consequence of the human not personally taking all actions in the system - of not being "in control" directly and immediately" (Miller & Parasuraman, 2007, p. 60).

Despite the small number of papers attributed to aviation, in explaining the problem of distancing, a large number of papers refer to the introduction of the glass cockpit as responsible for "psychologically distancing the pilots from their aircraft and the situation" (Stanton et al., 2001, p. 189). "'With old cockpits,' explained one pilot, 'the workload was high but you were always aware of what's going on.' In contrast, the glass cockpits are criticized for psychologically distancing pilots from their aircraft and environment" (Adams et al., 1995, p. 86). However, in discussing the phenomena of distance, probably the most obvious examples come from the domain of teleoperated vehicles or robots, as here the distance—again, both physically and cognitively—between work and the operator is the largest. "In the teleoperating environments, human perception is often compromised because the natural perceptual processing is decoupled from the physical environment" (Chen et al., 2007, pp. 1231-1232). As such, "one of the main challenges in achieving effective human-robot interaction will be bridging the cognitive gaps between these two entities" (Burke et al., 2004, p. 109).

The loss of manual skill, combined with the operators' detachment of the processes that they control, culminates in Bainbridge's 'ultimate' irony, which receives a lot of attention in the SA papers (mostly without explicitly referring to Bainbridge): As operators have become increasingly distanced from the processes they control, when in the case of an emergency the automation fails, they may have lost most of their (manual) skills to take over from the automation. "System operators working with automation have been found to have a diminished ability to detect system errors and subsequently perform tasks manually in the face of automation failures as compared with manual performance on the same tasks" (Endsley, 1995b, pp. 53-54). Moreover, the human operator—who has lost most of his manual skills due to his new role as a system monitor—is now supposed to take over from the automation during the most difficult and critical moment of operation (i.e. during an emergency): "This is where we come back to one of the pitfalls of early human–machine automation; the human-like substitute could fail at a critical point without leaving any choice to the human for regaining control of the situation …" (Tweedale et al., 2007, p. 1092).

The above listed automation-induced problems may be further exacerbated as the level of automation (LOA) increases. Various papers found "performance problems to be more significant under fully automated conditions than under intermediate LOAs" (Kaber et al., 2000, p. 412). Performance issues in our modern socio-technological systems are reported to be most problematic when automation takes operators 'out of the loop' (Sebok, 2000). The out-of-the-loop (OOTL) problems can be seen as the manifestation of the afore described problems that accompany the introduction of automation (Endsley & Kaber, 1999; Endsley & Kiris, 1995): "'Out-of-the-loop' (OOTL) performance may reduce a controller's ability to detect problems (e.g., conflicting aircraft), determine the current state of the system, understand what has happened and what courses of actions are needed, and react to the situation" (Kaber et al., 2006, p. 448). OOTL problems can be seen as a specific kind of distancing problem.

In short, when working with high levels of automation in the complex systems, many papers thus explain that it is increasingly difficult to be 'aware' of what the system is doing. “The out-of-the-loop performance can be linked to two major issues associated with the implementation of automation: loss of manual skills and loss of awareness of the state and processes of the system” (Endsley & Kiris, 1995, p. 381).

6.9 The discursive need for SA

The previous section has shown that the top cited papers on SA argue that operators working under high levels of automation in complex systems, run into various performance issues: information overload, workload problems, the distancing of the operator from real-time operations, which include out-of-the-loop performance problems. These issues are particularly problematic for the domains that make up the surfaces of emergence for SA, characterised by complexity, systems control tasks, and their safety criticality. Especially considering this last feature, the safety critical nature of these domains, there is paradoxically not a lot of room for any performance problems: “there is zero tolerance for errors in this application environment” (Kaber et al., 2000, p. 416). Consequently, the performance problems that accompany the widespread introduction of automation are a particularly salient issue for the domains and disciplines that invoke SA.

In addressing and specifying these critical and increasingly salient issues that accompany the introduction of automation, across the papers—across the domains and disciplines that invoke SA—the problems that arise in the various socio-technological systems are consistently attributed to the human component. This localisation of the problem in the human, rather than in the technology, can clearly be seen in the manner in which the automation-induced issues are presented: the human is struggling to cope with the vast amounts of information that automated systems produce; the human is encountering workload problems; the human is distanced from the machine; and, most evidently, the human is ‘out of the loop’—not the machine. Kaber et al. present a typical example:

> Specifically, operators may not detect critical system errors leading to automation failures; they may be inefficient in their responses to failures; they may lack awareness of system states and, consequently, knowledge of how to restore automated functioning due to absence from the direct control loop for extended periods of time. These performance problems not only affect

> productivity, but they can produce safety concerns as well. (Kaber et al., 2000, p. 410)

To stress the point: the human operator is not detecting something, is inefficient, and is lacking awareness and knowledge. This attribution is probably related to the finding that most papers (implicitly) regard the technological aspects of automation as highly reliable. Helmreich and Davies (1996, p. 278) share this observation: "Since this level of technological sophistication was introduced, failures have been related more to the weaknesses of humans than of machines".

Given this "increased realisation that system design is no longer optimised for human operation and, under some conditions, has 'overstepped the human's capability to keep track" (Stanton et al., 2001, p. 190), the SA papers thus start to focus on the question of how to optimise *human* performance under these new highly automated conditions. Wickens (2002, p. 131), for example, poses a question that is typical for many papers in the data set: "How to exploit the workload-reducing advantages of automation while still keeping the pilot adequately 'in the loop,' aware of actions taken by automated systems…"? In answering this question—in addressing the various issues that the human operator is having with highly automated technology—this is where the object of SA comes in: it provides a category—an object of knowledge—to shed light on the human part of the human-machine system. This is the discursive need that SA aims to fulfil.

In her seminal paper that tops the list of most cited SA papers, Endsley introduces the object of SA as a new—or rather, 'crucial'—manner of understanding the human performance problems that arise in complex highly automated systems.

> As technology has evolved, many complex, dynamic systems have been created that tax the abilities of humans to act as effective, timely decision makers when operating these systems. The operator's situation awareness (SA) will be presented as a crucial construct on which decision-making and performance in such systems hinge. (Endsley, 1995b, p. 32)

SA is presented as an object to better understand the problems that the human is having in modern highly automated systems. The object redefines the problems associated with the introduction of 'technology-centered' automation and focuses on designing automation that supports the human operator. The object emerges at the intersection of the social scientific authorities, as represented by human factors (and specifically health care), and those concerned with technology, as indicated by the computer sciences and robotics. "The value added by the SA concept is as a means of integrating these constructs in terms of the operator's overall goals and decision behavior. As such, this provides several advantages in the design process." (Endsley, 1995b, p. 51). A prominent design solution to come out of SA research, for example, is a concept that many papers in the data set refer to, that of 'adaptive automation': "The use of flexible and adaptive automation in supervisory control has been shown to promote improved automation monitoring and superior task performance as well as improved situation awareness in complex system management" (Cummings, 2004, p. 653).

From an archaeological perspective, however, it is important to stress that SA is not the result of this discursive need to address the social part of our socio-technological systems, but also constitutes and reinforces it. Archaeology is not concerned with epistemological questions of causation: the problem *is* not with the human; the problem *is* not with the technology; the problem *is* not even in the interaction of the human and technology. For Foucault, these are possibilities only because we have certain discursive practices that provide categories to understand the world in a particular manner. As Foucault explains (see section 2.1), the regularities we see in reality are created by the regularities of discourse, which we 'violently' impose on reality (Foucault, 1981). SA brings the human—specifically his or her awareness of the system—into view. With SA, when we have (technological) problems with systems, we can point to the human operator having problems with his awareness (SA) of the situation. This is the 'violence' that discourse does; this is what the object SA *does*, it imposes a particular meaning on reality.

Objects such as SA—just as other categories that provide psychological, human factors, or medical categories to understand the world—provide the discourse with a particular 'sensitivity' to human performance (problems): it makes the human part of the human-machine equation more susceptible for analysis. With SA, the human's awareness of the system is the problem, not the automation, not the system at large. The sensitivity on the human that SA produces is both a cause and effect of the influence of certain disciplines making its way into the (technical) fields of systems design and safety science. Archaeologically, it is thus not surprising that the 'psychology and human factors' and 'medical and information sciences' are such prevalent authorities of delimitation for SA. With SA, sciences that take the human as their central subject of analysis have something to say about these technical problems.

6.10 Conclusion

Inspired by Foucault's archaeological approach that sees objects of knowledge as contingent and historically constituted, this paper presents a study that set out to explore the emergence of SA as a ubiquitous scientific object. Examining the top cited papers on SA, it shows that the 'surfaces of emergence' of SA can be characterised by high levels of complexity, systems control tasks, and a central concern for safety. The 'authorities of delimitation' for SA, the scientific disciplines that invoke the object, are either concerned with SA in regards to humans, specifically in interaction with technology—the medical disciplines presented a special approach to this orientation—or they present a more technology-oriented approach to the study of SA. Across the papers, these social and technical disciplines report a number of interrelated concerns that humans—especially operators tasked with system control tasks in complex and safety critical systems—are having with the introduction of automation: information overload, workload problems, the distancing of the operator from real-time operations, and out-of-the-loop problems. In the literature, SA emerges as an object to address the need to better understand and design for the human part of these modern automated socio-technological systems.

SA does not just emerge to address these human performance problems, but also constitutes this attribution. SA is a category that discursively turns a human-machine problem into a human performance problem. It provides the various technological (and sociological) discourses that address these automation-induced issues with a particular 'sensitivity' to human performance problems. With SA, we are implicitly buying into the idea that humans are the fragile part of the system, rather than accepting that our technologies are not only ill-designed but also developing at a pace and in a manner that is undesirable—as evidenced by the problems with complexity, information overload, workload ironies, distancing of operators, and out-of-the-loop performance problems.

However, SA's intensive focus on humans can easily overstep its design goal and start to function differently. Object such as SA are so 'powerful'—that is, they provide such a strong focus on the human side of the human-machine system—that the context of their use may get lost. In some of our current discursive practices regarding SA, for example, we begin to believe that 'technology is reliable' and that the problem is thus entirely with the human. SA is used as an endogenous—that is, internal to humans—explanation for human error: poor SA is a personality issue, a motivational problem of people not trying hard enough, or due to people's inherent performance limitations and cognitive shortcomings. The object's sensitivity for the human part of the system turns SA into a stand-in for the object of human error, rather than an explanation for how automated systems place human operators in a difficult configuration. The epistemological problems relating to this have been extensively discussed in the human factors literature (see, for example, Billings, 1995; Dekker & Hollnagel, 2004; Flach, 1995b).

The value of Foucault's archaeological approach lies in describing our discursive practices and the effects that they may have. SA is not the necessary culmination of scientific progress, but rather a contingent effect of some historical practices and 'conditions of possibility' (see Winsen, Henriqson, Schuler, & Dekker, 2015). As an emancipatory project, archaeology allows us to ask new questions. Foucault asks, for example, why do we constitute the world in one way rather than another? Why do these papers state that the 'human is out of the loop'; why not say that 'technology is out of the loop'? Instead of saying that the operator lost 'situational awareness', why do we not say that the operating environment has acquired 'situational opacity'; that the environment is no longer communicating relevant elements to the human agent? In the latter case, the problem, and the locus of intervention for improving the human-machine system, is the system rather than the human.

Chapter 7:
Situation Awareness: Some Conditions of Possibility

Statement of Contribution to Co-Authored Paper

This chapter includes a co-authored paper. The copyright of this article has been transferred to Taylor & Francis, however, authors have the right to include their article in a thesis or dissertation that is not to be published commercially.

The bibliographic details of the co-authored paper, including all authors, are:

Winsen, R. van, Henriqson, E., Schuler, B., & Dekker, S. W. A. (2015). Situation Awareness: Some Conditions of Possibility. *Theoretical Issues in Ergonomics Science, 16,* 53-68.

As a the first-author of this paper I have had a leading role in pursuing this argument; I have reviewed the literature on SA; I have designed the theoretical and analytical approach of this paper; I have engaged the co-authors in discussions about the arguments pursued in the paper; I have implemented the feedback from the co-authors; I have written all the text that makes up the final version of the paper; I have critically revised numerous versions of the paper; I have implemented the (minor) feedback from the reviewers.

(Signed) ______________________ (Date) 16-02-2015
Roel van Winsen

(Countersigned) ______________________ (Date) ___________
Sidney W. A. Dekker (Principal supervisor and co-author)

Situation Awareness: Some Conditions of Possibility

Abstract

Situation Awareness (SA) has become a ubiquitous object of knowledge in our discourses of human performance and accident explanation. Based on Michel Foucault's archaeological approach, in this paper we examine SA by mapping the 'conditions of possibility' for this object to emerge. By highlighting the logic that SA builds upon, the political need that it intends to address, and the knowledges that delimitate it in its constitution, we aim to display the contingent nature of this object. Ultimately, we argue that as a discursive object, SA has effects.

7.1 Introduction

"The flight crew lost situational awareness and commenced takeoff from the wrong runway" (Aviation Safety Council, 2002, p. 237). This was the 'probable cause' coming out of the investigation of the Singapore 006 accident in Taiwan in 2000, but could have been from almost any (aviation) accident investigation report produced over the last decade. Situation awareness—and the loss of it—is everywhere; there are books, scientific journals, conferences, training courses, operating manuals, design guidelines, and accidents investigation reports addressing it. Nowadays there are even people held criminally negligible for losing it. Typing 'situation awareness' into Google provides more than 500,000 hits, and these results do not even take into account variations of the term such as 'situational awareness' or 'SA'. Situation awareness is a ubiquitous object of knowledge in our discourses of human performance and accident explanation.

Objects of knowledge, such as situation awareness (hereafter referred to as SA) are the contingent outcome of certain historical discursive practices. Michel Foucault's archaeology allows for a comprehensive examination of discourses to understand how particular conceptions of the world become fixed and pass as truth. The goal of an archaeological approach is not to make any judgements about the validity of discourses and its claims, but rather to investigate the 'conditions of possibility' that govern the production of discourses and objects in it.

In this study we examine SA as an object produced by a complex discursive system. This paper is not a discussion about the usefulness or ontological status of SA as an object. Rather it is about mapping how—under what conditions of possibility— SA emerged and functions as a universal object in the human factors and safety discourses. We intend to highlight the contingent nature of this object by describing the logic that it builds upon, the political need that it intends to address, and the knowledges that delimitate it in its constitution. In conclusion, we argue that discursive objects, such as SA, have effects.

7.2 Archaeology, Objects, and Conditions of Possibility

Every mode of thinking involves some implicit rules that restrict the range of possible discourses within a certain time and place (Foucault, 2002a). In his archaeological projects, Foucault argues against seeing the objects of our (scientific) knowledges as natural or necessary—that is, as the logical endpoint of (scientific) progress—but rather as constituted by historically contingent discursive practices. Archaeology is "an inquiry whose aim is to rediscover on what basis knowledge and theory became possible; within what space of order knowledge was constituted; on the basis of what historical *a priori* … ideas could appear" (Foucault, 2002b, pp. xxi-xxii). Discourses, and objects in them, have a history, rather than a transcendental truth (Foucault, 2002a). Through understanding how particular conceptions of the world emerge, we can see the arbitrariness of our discourses and objects. It opens up possibilities for the examination of taken-for-granted truths.

Social science disciplines—like psychology, sociology, and we will argue the same holds for human factors and safety science—suffer from a retrospective illusion, which prevents them from seeing that their objects are historically constituted (Rose, N. S., 1996). Their objects are not 'out there' in the world waiting to be 'discovered'. As an example, Foucault explains that the "psychiatric discourse finds a way of limiting its domain, of defining what it is talking about, of giving it the status of an object – and therefore of making it manifest, nameable, and describable" (Foucault, 2002a, p. 46). Objects are the effects of certain discursive practices and do not rely on a reference to an *a priori* reality (Foucault, 2002a). From an archaeological perspective it is thus apt to examine the 'conditions of possibility' for the conception of social-science objects, such as SA.

With conditions of possibility, Foucault means the:

> … conditions necessary for the appearance of an object of discourse, the historical conditions required if one is to 'say anything' about it, and if several people are to say different things about it, the conditions necessary if it is to exist in relation to other objects, if it is to establish with them relations of resemblance, proximity, distance, difference, transformation – as we see, these conditions are many and imposing. Which means that one cannot speak of anything at any time; it is not easy to say something new; it is not enough for us to open our eyes, to pay attention, or to be aware, for new objects suddenly lo light up and emerge out of the ground. (Foucault, 2002a, p. 49)

New scientific objects, such as SA, don't 'suddenly emerge out of the ground', but are already tacit in the knowledges on which they finally find their manifest form. The conditions of possibility for SA thus address the particular conditions under which SA has been allowed 'in the true' (Foucault, 1981; Foucault, 2002a; Rose, N. S., 1996).

In this study we sketch some conditions of possibility for SA to emerge as an object of the human factors and safety discourse. The conditions of possibility for any discourse or object should not be seen as separate and autonomous, as they constitute a complex and contingent network of discursive practices that mutually constitute and reinforce each other. However, for analytical reasons, after describing the object of our analysis, we separate the logic, the political need, and the knowledges that make up the conditions of possibility for the emergence of SA.

7.3. The Object of SA

Originally adapted from its more commonsensical use in the air force—SA was named as early as World War I by Oswald Boelke who realised "the importance of gaining an awareness of the enemy before the enemy gained a similar awareness" (Endsley, 1988, p. 97)—the concept of SA was picked up by the human factors and safety communities during the late 1980's. However, it was not until the journal of Human Factors published its special issue on situation awareness in 1995 that the concept really exploded as a scientific object (Patrick & Morgan, 2010). Even though the usefulness and validity of SA as a scientific construct is still fervently debated, it is nowadays an important object in the scientific disciplines of human factors and safety science. SA has been investigated in various contexts, including aviation (Jones & Endsley, 1996; Salas & Dietz, 2011), health care (Gaba et al., 1995; Schulz et al., 2013), driving (Horswill & McKenna, 2004; Ma & Kaber, 2007), nuclear power (Patrick & Morgan, 2010), energy distribution (Salmon et al., 2007), and sport (James & Patrick, 2004), to name just a few.

In addition to the plethora of scientific discourses in which SA surfaced, its conception and modelling was accompanied by various ways of measuring it in practice (for a comparison of three popular techniques, see Salmon et al., 2009). This resulted in widespread rise of training programmes—as such, SA has become an integral part of CRM training—and assessment of individual operators and teams for SA (Robertson & Endsley, 1997). We also see SA surface in the design of human-machine interfaces (Endsley, Bolte, & Jones, 2003; Jin & Wang, 2009). Additionally, 'losing situation awareness' or 'poor situation awareness' have become legitimate explanations for human error. Stanton, Chambers and Piggott (2001, p. 190), for example, report that "[o]ne review of over 200 aircraft accidents found that poor situational awareness was the main causal factor." Subsequently, the construct has spread into the judicial discourse, as we now see people held criminally negligent for losing SA (for example, Supreme Court of British Columbia, 2013).

The application of SA throughout the vast range of academic disciplines and other institutions makes it difficult to understand and uniformly define the construct. To circumvent this problem, as well as a sign of its frequent use and status, some researchers leave it undefined, assuming that SA—an awareness of the situation—is a common-sensical notion. When it is explicitly defined, most studies refer to Endley's three-stage description of SA as: "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (1988, p. 97). In her seminal articles on the theory (1995b) and measurement (1995a) of SA, Endsley presents SA as a "crucial construct on which decision making and performance in such systems hinge" (1995b, p. 32).

Endley's hierarchical model promises to better explain (poor) human performance and provide a new approach to the design of human-machine interfaces by distinguishing between three levels of SA, in which the higher levels of SA depend on the success of the lower levels. The first step in achieving SA, or 'Level 1 SA', is to perceive the relevant elements in the dynamic environment. 'Level 2 SA' constitutes the formation of a holistic picture of the environment by comprehending the meaning and significance of these disjointed elements in light of the operator's goals. Finally, 'Level 3 SA' is about predicting future states of the environment, by projecting the actions of the elements into the future .

Even though Endley's (1995b) well-ordered structural model easily crosses domains and provides vast intuitive appeal, there is still little consensus on the theoretical and methodological approaches that have been used to study SA. Over the years, a plethora of other definitions and specifications for modelling SA have been proposed. Besides being conceptualised as the product of information processing (Endsley, 1995b), SA has been referred to as cognitive management—"the up-to-the-minute cognizance required to operate or maintain a system"—(Adams et al., 1995, p. 85), as the process of dynamically directing attention and assessment (Gaba et al., 1995), as "a label for a variety of cognitive processing activities" (Sarter & Woods, 1995, p. 16), as adaptive, externally directed consciousness (Smith & Hancock, 1995), or even as a search for meaning within complex work domains (Flach, 1996).

While different, these definitions do share to a large extent the same theoretical underpinnings. The largest divergence pivots around the question whether SA is the cognitive product of awareness (Endsley, 1995b) or the process of gaining awareness (Durso & Gronlund, 1999; Salmon et al., 2007; Tenney & Pew, 2006). "In this distinction, product refers to the state of awareness with respect to information and knowledge, whereas process refers to the various perceptual and cognitive activities involved in constructing, updating, and revising the state of awareness" (Adams et al., 1995, p. 88). Conceptions of SA as a state often build their model of SA on a cognitive theory that uses an information processing approach, whereas SA as a process, sometimes referred to as 'situation assessment', builds on an ecological cognitive approach underpinned by Neisser's (1976) perceptual cycle.

The various ways of conceptualising SA, either as a cognitive state or process, help to constitute and reinforce SA as an object in the analysis of human performance. As Salas and Dietz explain: "Despite the absence of a unifying definition, theoretical issues surrounding SA coalesce to provide a foundation for its understanding (2011, p. xv)". While at the surface these different conceptions of SA may seem vast, at an archaeological level they all refer to the same discursive object, as it can be captured under the same conditions of possibility.

7.4 Condition of Possibility: The Logic of Representation

For SA to emerge as a possible object we need a conception of the subject and reality that allows for an internal representation of an external material world—we need a particular rationality: "This idea of separation between the human operators' [sic] understanding of system status and actual system status is at the crux of the definition of situational awareness" (Stanton et al., 2001, p. 189). According to Foucault (1981), around the time of Plato a certain division was introduced that separated true discourse from false discourse. This division gave rise to our 'will to truth', the will to produce true representations; to present accurate (true) 'copies' (using Plato's vocabulary) of our external reality. Since this Platonic division, we take this concern with representing the truth as necessary and self-evident. According to this 'logic of representation', there is a transcendental reality (i.e. meaning in the world), which can be discovered and represented. This rationality also endows humans with an *a priori* interiority, which can represent this external reality. Modern philosophy and science are also based on this logic, assuming a transcendental truth that can be accessed and represented by using the proper (scientific) methods.

In our modern western conception of the person, the self is construed as a naturally discrete entity, for which the boundaries of the body enclose an inner life of the psyche (Rose, N. S., 1996). Since the late 18th century, we started to understand the world through psychological categories such as interiority, consciousness, and identity, based on general laws of 'human nature.' Gergen points to the enlightenment as the birthplace of this modern conception of the individual endowed with an interiority, a consciousness: "Most particularly [the enlightenment] gave intelligibility to the idea that each individual is capable of observing the world for what it is … that is, the capacities to observe for oneself, to think, evaluate, and then to choose one's actions" (1999, p. 7).

Such a conception of the self gives rise to a dualist ontology, which assumes a Cartesian divide between a psychological inner world—a consciousness—and a material external world. “Virtually all theories of SA rely on the idea of correspondence – a match, or correlation, between an external world of stimuli and an internal world of mental representations” (Dekker & Lutzhoft, 2004, p. 22). Moreover, not only does SA build on this Cartesian separation of mind and matter, it also assumes that the mind is capable of accurately representing this external world somehow. “Strong correspondence between the interpretation and the objective situation means high situation awareness. Weak correspondence means low situation awareness” (Flach, 1996, p. 3). The mind is supposed to work as a ‘mirror’ representing the world (Dekker, S. W. A. et al., 2010).

SA thus builds on a ‘logic of representation’ (Foucault, 2002a, 2002b), also referred to as the ‘correspondence theory of truth’ (Willig, 2008), as a condition of possibility. Such rationality suggests that there is a correspondence between the external world—objects, events, phenomena—and our internal understanding of it. However, this kind of representation does not only take place within the individual, when an operator has some degree of SA (has a correct or incomplete representation of the world, that is, good or poor SA), but also at the level of SA as a describable and measurable scientific object, which is supposed to represent a psychological reality.

Our modern conception of the individual self is closely aligned with assumptions of truth and science (Gergen, 1999). The dualist ontology and the modern conception of the individual, contributes to the belief, first, that there is an *a priori* reality to be represented, and second, that there is only one correct representation of that reality. This conception of the individual as being able to objectively represent the external world is where SA as a scientific object stands on, as now it becomes a matter of assessing the accuracy of the people’s representations in regards to the true state of the world (Dekker, 2012a). “SA represents a continuous diagnosis of the state of a dynamic world. As such, there is a ‘ground truth’ against which its accuracy can be assessed (e.g., the objective state of the world or the objective unfolding of events that are predicted)” (Parasuraman et al., 2008, p. 144).

Building on the logic of representation, the scientific object of SA becomes possible, as we can now ask questions about the nature and accuracy of people’s representation. We can develop theories, devices, and experiments to specify and measure it. With this, we can quantify the accuracy of people's representations and, depending on the specificity of the model, we can locate people's 'errors' in obtaining this representation. For example, in investigating the type of SA errors in aviation, Jones and Endsley (1996) found that found that 76.3% of the errors were Level 1 SA (failure to correctly perceive information), 20.3% were Level 2 SA (failure to comprehend information), and 3.4% were Level 3 SA (failure to project information into the future). By providing a proper method to access reality, SA as a scientific object, produces a predictable and normative reality.

The logic of representation is thus a necessary condition of possibility for SA to emerge as a scientific object. This rationality does not only endow the individual with an internal self, possible of representing an external world, but also enables SA as a describable and measurable scientific object (that represents a transcendental reality). Without such rationality, SA could not have emerged as a meaningful object.

7.5. Condition of Possibility: The Political Need to Address Human Performance in Technological Systems

The logic of representation enables the scientific study of people's internal awareness of an external situation, however, this rationality does not explain the emergence of SA as a possible scientific object. We need to have a particular need to invoke this object in our (scientific) explanations of the world. In other words, we need particular 'events' (Foucault, 2002a) for which the object will provide us with new leverage, new insights, new solutions. In this section we will argue that the second condition of possibility for the emergence of SA is the political need to address the increasingly conspicuous problem of human performance in our socio-technological systems. This political need is entwined with, at least, two important historical issues: the increasing complexity and opaqueness of our modern technologies and the occurrence of a number of politically relevant accidents.

The 20th century can be characterized by the rapid development of technological systems. From the simple mechanical 'amplification' of human physical strength emerged a new type of tools, aimed not only at physical tasks, but also at improving the efficiency and scope of perceptual and cognitive tasks. "Amplification can therefore be seen as a way of overcoming the limitations of the unaided human – whether these refer to physiological, biomechanical, perceptual, or intellectual (mental or cognitive) functions" (Hollnagel & Woods, 2005, p. 25). As mechanisation (as a manner to amplify human muscular power) has been rapidly followed by automation (systems being controlled by autonomous machines rather than humans), centralization (controlling various systems from a single location), and computerization of our systems (Hollnagel & Woods, 2005), we currently live in a 'technologically intensive' age (Saleh & Marais, 2006).

Amplification comes with a price. Consisting of an ever-increasing number of parts and interactions, our technological systems become increasingly complex, emergent, and non-linear (see Dekker 2011). Amplification also increases the distance between the human and the operations that they are responsible for. Particularly in the case of automation, computers mediate the interaction between humans and the current state of the system. As an example, "[m]any researchers cite the glass cockpit [the replacement of the multitude of dials in a traditional cockpit with a small number of digital displays that need to be accessed sequentially] … as responsible for psychologically distancing the pilots from their aircraft and the situation" (Stanton et al., 2001, p. 189).

The increasing complexity, as well as the detachment of the operator from real-time operations, leads to an increasing opaqueness of our socio-technological systems. This, in turn, might lead to 'out-of-the-loop performance' (Endsley & Kaber, 1999). As our systems become increasingly opaque, it gets harder for operators working with and within them to remain (situationally) aware of the system's current state. "'With old cockpits,' explained one pilot, 'the workload was high but you were always aware of what's going on.'" (Adams et al., 1995, p. 86). In this sense, software has added greatly to the opaqueness of our systems: "With software, the possible states that a system can end up in becomes mind-boggling" (Dekker, S. W. A., 2011, p. 196).

Also adding to the opaqueness of our automated systems is the vexing amount of information that they can produce, "both on the status of their own components and on the status of the external environment. … The problem with today's systems is not a lack of information, but finding what is needed when it is needed" (Endsley, 2000, p. 4). This particular ‘data overload’ problem has been termed the ‘data availability paradox’, as on the one hand we have more and more data available, but on the other hand we realize that this abundance of data makes it increasingly difficult to find meaningful information in the light of our goals and tasks (Woods, Patterson, & Roth, 2002).

With the rapid growth of technology, the tasks, roles, and challenges of the human in our socio-technological systems have changed. The changing nature of our socio-technological systems, comes with “opportunities for new kinds of system failure that did not exist in older, simpler systems” (Cook & Woods, 1996, p. 594). This change, however, has for the majority of the 20th century not been matched by a change in the strategy to control our safety-critical systems. The practical management and the scientific study of safety in the twentieth century mainly focused on increasing the technical reliability of our socio-technological systems. “When therefore an accident – and in particular a spectacular accident – happened there was a natural tendency to look for causes that could be expressed in terms of technological failures” (Hollnagel, 2004, p. 29). Safety, as the absence of accidents, was assumed to be achieved through ensuring technical reliability, which was often operationalised as technical redundancy, achieved through either duplication or overlap of parts.

The increased complexity and opaqueness of our socio-technological systems was particularly exemplified by a number of catastrophic accidents during the late 1970’s and 1980’s that turned safety into a topic of social—that is, political— interest. Accidents such as the Tenerife Airport disaster (1977), the Three Mile Island nuclear incident (1979), the Bophal chemical accident (1984), the space shuttle Challenger blow-up (1986), the Chernobyl nuclear accident (1986), the capsizing of the MS Herald of Free Enterprise (1987), London’s King’s Cross station fire (1987), the Piper Alpha rig blowout (1988), and the Clapham Junction train crash (1988) dramatically showed the limits of improving safety through improving the reliability of the technical parts of our systems.

These accidents showed that highly automated systems with multiple layers of redundancy—systems that were assumed to be fail-safe—were still susceptible to failure. As such, the events in the late 1970’s and 1980’s clearly displayed safety science’ inability to understand and control our increasingly complex and opaque safety-critical technologies. Public demand for explanations of these dramatic events triggered a vast amount of research into the complexities of accidents, with an increasing focus on the human and the social. Human performance in our socio-technological systems became a more salient issue.

The political need to address human performance— explain the social contribution to accidents and better design human-machine interfaces— in our increasingly complex and opaque socio-technological systems constitutes a second condition of possibility for the emergence of SA as a scientific object. This need can explicitly be traced in justifications of SA research. Salas and Dietz explain in the introduction to their book on situation awareness in aviation, that “the technological advances designed to improve safety and efficiency also increased system complexity, thereby engendering parallel challenges for optimizing human ”, as a result “research related to SA burgeoned in the 1980s to address human performance issues associated with technological advances in the cockpit” (2011, p. xiv). Similarly, Endsley explains in her seminal 1995 paper, that “[a]s technology has evolved, many complex, dynamic systems have been created that tax the abilities of humans ... The operator’s situation awareness (SA) will be presented as a crucial construct on which decision making and performance in such systems hinge” (Endsley, 1995b, p. 32).

Just like the logic of representation as a first condition of possibility, SA shares this second condition of possibility with many other scientific objects, especially those in the safety and human factors domains. The birth of scientific objects such as Human Reliability Assessment (HRA), Crew Resource Management (CRM), and Safety Culture can all be related to this same political need. In historicising HRA, for example, Hollnagel (2005, p. 160) shows “the strong connection between the accident at Three-Mile Island (TMI) on March 28, 1979, and the growth in the number of HRA methods.” Similarly, we can link the conception of the objects CRM and safety culture to, respectively, the Tenerife airport disaster and the Chernobyl nuclear accident. Having said this, an archaeological perspective is not concerned with finding the historical origin of these objects, but rather with understanding the particular discursive need out of which they emerged. With the saliency of human performance problems in our complex safety-critical systems, there is a strong impetus to explain human error, fix the human operator in our safety-critical systems, and better design human-machine interfaces. SA is an attempt to do so.

In summary, with technology changing the nature of our work and accidents came the political need to address the conspicuous issue of human performance in our increasingly opaque and complex socio-technological systems. This political need, to better explain the social contribution of accidents and prevent them through better training and (interface) design, was fuelled by the large industrial accidents of the late 1970s and 1980s and constitutes the second condition of possibility for the emergence of SA as a scientific object.

7.6 Condition of Possibility: the Cognitive Knowledges

The logic of representation and the political need to address the problematic human component of our increasingly complex and opaque safety-critical systems are two conditions of possibility for the emergence of SA as a possible scientific object. As a third condition, we need particular knowledges in which, and upon which, this object can emerge (Foucault, 2002a). In this section, we will argue that the cognitive knowledges (disciplines as cognitive psychology and human factors) provide a substratum of knowledge—a 'menu' of scientific theories, prepositions, and objects—for SA to take-up and manifest upon.

In addressing the political need to fix the conspicuous issue of the human in our technological system, whom do we turn to? What discipline is concerned with the study of humans, or more specifically, with the study of humans in interaction with technology? Psychologists and human factors specialists, respectively. When Descartes suggested that mind and matter are separate, he also argued that these two realms should be studied separately (Dekker, 2005), effectively constituting the need for a science of the mind: psychology. As the scientific study of consciousness or subjective mental life, Descartes and later Wundt, argued that the realm of the mind can be accessed through psychological methods such as introspection. However, when at the turn of the 20th century the subjective method of introspection was deemed as unscientific, it was superseded by behaviourism, as an anti-subjective approach to psychology (Dekker, 2005; Smith & Hancock, 1995; Suchman, 2007).

According to behaviourism, "psychology must discard all reference to consciousness" (Watson, 1913, p. 163) as psychology should take behaviour as its object of study and needs to rely on "direct observation under experimental conditions" (p. 161) as its method. What goes on in the mind of the human should not be the object of psychological research, as his is something we cannot get to—a 'black box'. In studying behaviour (mostly that of animals) in an observable and systematic manner, behaviourist devised all sorts of laboratory experiments to understand stimulus-response behaviours acquired through 'conditioning'—a method of learning that builds on rewards and punishments of behaviour. The price that behaviourism pays for staying away from 'mentalist' explanations of behaviour was to "confine experimental psychology to relatively simple memory and learning experiments, and to a preoccupation with laboratory rats rather than humans engaged in complex thinking and problem-solving task" (Simon, 1980, p. 76).

Behaviourism works well for understanding simple human (stimulus-response) behaviour, particularly in controlled laboratory settings, however, for more complex behaviour it becomes problematic. With World War II bringing a new level of complexity to our socio-technological systems, the nature of human work changed. Behaviourism's principle to only study overt behaviour no longer seemed a viable way to understand human performance. "No matter how clever a system of rewards and punishments psychologists set up, radar operators monitoring for German aircraft intruding into Britain across the Channel would still lose their vigilance over time" (Dekker, 2005, p. 105). With the increasing opaqueness and complexity of human-machine interactions, the basic assumption that human performance could be described without making any references to what is happening inside the 'black box' became increasingly difficult. Psychologists needed different explanations than the ones that behaviourism was able to offer. In short, they needed to open the 'black box'.

In order to deal with more complex human behaviour, the "cognitive revolution reintroduced 'mind' as a legitimate object of study. Rather than manipulating the effect of stimuli on overt responses, it concerned itself with "meaning" as the central concept of psychology" (Dekker, 2005, p. 105). Kick-starting the cognitive revolution, the cybernetics movement began during World War II, when a group of mathematicians, neuroscientists, and engineers tried to create a formal, mathematic 'science of mind'. By taking neurons as the basic elements of the brain, cybernetic studies tried to provide a description of the brain in formal logical terms. This approach turned out to be critical for the invention of computers, which in turn provided the conceptual basis for a new scientific study of mind (Capra, 1996).

> The computer model of mental activity became the prevalent view of cognitive science and dominated all brain research for the next thirty years. The basic idea was that human intelligence resembles that of a computer to such an extent that cognition – the process of knowing – can be defined as information processing, i.e. as manipulation of symbols based on a set of rules. (Capra, 1996, p. 66)

Building on the Shannon-Weaver model from information theory, cognitive psychology now aimed to describe the internal processes of the mind as a series of transformations of information (Hollnagel & Woods, 2005). Information processing models understand cognition as a system that receives information, represents it symbolically, and transforms these representations in various stages. As such, the information processing approach, with its computational metaphor, provides "a detailed analysis and specification of psychological activities in terms of component processes and procedures" (Barber 1988 p. 19). In outlining these various component processes, the computer metaphor provided the "recipe for selecting, storing, recovering, combining, outputting and generally manipulating information" (Neisser, 1976, p. 8). Humans, conceived as information processors, were now endowed with a vast number of cognitive 'components', such as attention, perception, decision-making, (sensory, short and long-tem) memory, problem solving, action selection, and many more.

The birth of human factors as a multidisciplinary science, concerned with the study of the interaction of humans and technology, largely coincided with the cognitive revolution (Meister, 1999). Human factors researchers introduced experimental psychology's methods and theories to the study of human-machine interactions (Coury et al., 2010). As such, it took the information-processing objects of psychology—attention, perception, memory, etc.—into the realms of the design, analysis, management, and maintenance of complex socio-technological systems. "Today, human factors have a sizeable stock of concepts that are used to express insights about the functional characteristics of the human mind – or covert information processes – that underlie complex behaviour" (Dekker & Hollnagel, 2004, p. 79).

In accident reports from the 1960's and 1970's we see human factors specialists bring the mentalist language of cognitive psychology, mainly that of information processing, to explanations of (poor) human performance and accidents. Accident reports are rich in explanations as: ‘failure to monitor,’ ‘faulty’ or ‘poor judgment,’ ‘failure to take action,’ ‘distraction of attention,’ ‘spatial disorientation,’ and ‘pilot’s misinterpretation,’ to name just a few (Ministry of Transport, 1968; National Transportation Safety Board, 1974; 1976. NB. The three accident reports analysed here would nowadays probably have substituted these explanations for the label ‘loss of SA’.). As these ‘explanations’ are all indications of problems that operators are having with technology, and given the political need to better understand human performance and prevent accidents, the next step for cognitive psychologists and human factors researchers was to refine the mechanisms underlying these ‘labels’.

In further opening the ‘black box,’ the “interdependence of memory, perception, and action has been the focus of much cognitive theory and research, including basic experimental work, modelling efforts in the neural network domain, and work on plan and goal structures from the computing and engineering disciplines” (Adams et al., 1995, p. 88). As a result, the human factors discipline saw a wave of new models that aimed to integrate various mentalist elements to better understand complex human performance. “[T]he concept of consciousness has enjoyed a gradual rehabilitation, albeit cloaked in many guises. Constructs such as attention, mental workload, and now situation awareness (SA) have arisen for consideration” (Smith & Hancock, 1995, p. 137). Amongst other attempts, SA thus presents such a specification of the deeper mechanisms making up human cognition (another prime example is 'sensemaking,' see Weick, 1993).

SA systematically addresses the relation between various cognitive elements. “SA presents a level of focus that goes beyond traditional information processing approaches in attempting to explain human behaviour in operating complex systems” (Endsley, 1995b, p. 32). Whether the object is conceptualized from an information processing perspective—as the product of the interactions of a number of cognitive component (see Endsley, 1995b)—or from a more ecological perspective—building on Neisser’s (1976) perceptual cycle to explain how perceptual exploration of the environment modifies the schemata that in turn direct this perceptual exploration (see Adams et al., 1995)—all models of SA aim to integrate a number of cognitive elements into a single mechanism. As such, SA is an object that opens the ‘black box’ by integrating various cognitive elements to better understand the mechanisms that underlie complex human performance (particularly in interaction with complex dynamic technological systems).

The cognitive revolution and its resulting knowledges can thus be seen as a third condition of possibility for the emergence of SA as a scientific object. Within the behaviourist paradigm of psychology the object of SA would not have been able to emerge. Building on the other conditions of possibility, the mentalist psychological knowledges provide a 'menu' of cognitive elements and theoretical perspectives for SA to take up and manifest upon.

7.7 Conclusion

Once we have a particular logic, need, and knowledges, we see the object of our analysis emerge. SA is the product of the logic of representation, the political need to understand complex human behaviour in our socio-technological systems, articulated upon and specified by cognitive discourses of psychology and human factors. With SA, we now have an object that aims to better explain the cognition underlying complex human behaviour, and as such, helps with the design of human-machine interfaces and provides new insight into the occurrences of accidents.

Our archaeological analysis of the constitution of SA shows that this is not a natural object, but rather a historically contingent object—an artefact of discursive practices. The object is an effect of certain conditions of possibility, historically constituted through certain discursive practices. Describing this object as 'contingent' means that the historical emergence of it is not some logical end-point of scientific progress, but merely one possible outcome of a complex network of relations between historical events, rationalities, and contemporary discourses. Foucault's archaeological approach is thus explicitly anti-teleological, as it does not involve any assumptions of progress (nor regress!).

Archaeologically, the discussion about whether SA is the cognitive product of awareness or the cognitive processes of gaining awareness is just a matter of delimitating the object differently. Both conceptions are different delimitations of the same discursive object, as they function under the same conditions of possibility—they address the same need, are based on same rationality, and largely draw on a similar substratum of knowledge.

Similarly, the questioning—on-going since SA's first scientific conception (see Dekker, 2012a; Dekker, S. W. A. et al., 2010; Flach, 1995b; Sarter & Woods, 1991)—of the epistemology and usefulness of this object is not of archaeological concern. Whether SA is a cognitive state or a process, whether it is real or not, or even whether it is useful or not, does not matter from an archaeological point of view. Archaeology is not interested in the epistemological or ontological status of the object. This does not mean that, whilst outlining certain discursive practices as conditions of possibility, archaeology does not include certain ontological or epistemological assumptions as conditions that enable these discursive practices. For example, we identified the logic of representation—a positivist epistemology—as a condition of possibility for SA to emerge as a modern scientific object. The point is, however, that archaeology does not take a particular stance on the epistemological or ontological status of the object. What matters from an archaeological perspective is that this is an object at the level of the discourse.

Once the object is constituted, given its conditions of possibility, a number of (self-correcting and self-reinforcing) discursive practices start defining, delimiting, and correcting it. Various institutions provide a number of specifications for modelling the object and legitimising the further use of it. This is done through various 'enunciative modalities' (Foucault, 2002a) such as scientific research, assessment methods, training programs, statistical records, legislation, design recommendations, accident investigation reports, managerial reproaches, and recently even through court verdicts, to name just a few of SA's materialities. The use of the object in these kind of discursive practices reinforces its status—the more it is used, the greater the consensual authority on using it. Even this very paper is contributing to its status as an object.

The Singapore 006 accident investigation (Aviation Safety Council, 2002) provides an example of the powerful manner in which the object of SA is used and reinforced. Not only does the report refer to SA, and the loss of it, more than a couple dozen times as one of the contributing factors of the accident, but the object also surfaces in one of the 'safety actions' that have been implemented in the aftermath of the accident: "A new CRM training programme for pilots has been developed and implemented, which includes situational awareness and error management training as separate modules" (p.

249). This kind of discursive use of the object contributes to produce and reinforce its status among other objects in the safety discourse.

As a critique to these kind of investigatory conclusions and recommendations, a number of researchers have questioned the causal power that has been ascribed to SA, saying that 'constructs don't cause anything' (Dekker & Hollnagel, 2004; Flach, 1995b): "We heard here that deficient SA was a causal factor in many airline accidents associated with human error. We must avoid this trap: deficient situation awareness doesn't cause anything" (Billing, 1996, as read in Dekker & Hollnagel, 2004, p. 80). Again, from an archaeological perspective it does not matter whether the object provides the necessary specification of the mechanism responsible for such causation, it does not concern itself with such epistemological questions. What is clear from an archaeological position, however, is that constructs—objects such as SA—have (causal) effects: accidents are attributed to a loss of SA, corresponding training programs are set up, procedures are revised, operators are being criminalized, and people end up in jail for losing SA.

From an archaeological perspective, language does not represent reality, but constitutes it (Foucault, 2002a). Words do not merely name certain phenomena, but they produce them—they produce meaning through particular political 'regimes of truth' (Foucault, 1980). As artefacts of discursive practices, objects set up an entire agenda of institutional practices. With SA, we constitute a set of institutional techniques (aimed, for example, at improving safety). We create performance assessment measures that are extremely powerful to subjectivate people in a certain manner. We create certain moralities based on this object. Objects are the effects of certain discursive practices and, in turn, objects have certain effects.

As an example: Even though the accident investigation report does not explicitly lists SA as a direct cause of the sinking of the passenger ferry Queen of the North, it does state that, "unsafe navigation practices persisted that, in this occurrence, contributed to the loss of situational awareness by the bridge team" (Transportation Safety Board of Canada, 2008, p. 40). In the subsequent court case against the commanding officer, Karl Lilgert, the final verdict draws on the object of SA to make the case for the accused's 'dereliction of his duty'—which resulted in his conviction for criminal negligence:

> Maintaining situational awareness at all times and in all circumstances is key to proper navigation. The evidence of the expert, Mr. Flotre, was Mr. Lilgert failed to comply with the collision regulations by not standing watch using his eyes, ears, the ECS, and radar. (Supreme Court of British Columbia, 2013, p. 6)

In short (whilst deliberately simplifying), Karl Lilgert is in jail for losing SA; objects have effects. The objects that we constitute in particular discourses for particular purposes—such as SA as an object to help explain complex behaviour in human-machine interactions—are borrowed by other discursive practices and used for different purposes—for example as a convenient and holistic construct to explain criminally negligent behaviour. Future research might provide answers as to how certain objects are appropriated by other discourses. How human factors objects, such SA, help to constitute a discourse on criminalisation. Or, vice versa, how criminality—with juridical objects as negligence, guilt, and recklessness—has become part of the human factors and safety discourses.

SA is thus not only a scientific object, to be discussed and analysed in terms of its specifications, its usefulness, and the methodologies to investigate it. By arguing for the contingent nature of our objects of knowledge, archaeology opens another agenda of inquiry. Archaeology opens an ethical discussion, asking questions about the consequences for operating a particular discursive object. Discursive objects that the human factors community once introduced to help operators better understand and deal with complex dynamic environments are now turned against them as they provide normative standards for behaviour. The contribution of this paper is not in making an argument about whether SA is real—or even useful—or not. It does not take a position on whether we are researching SA in an appropriate manner or not. Rather, we aim to argue that it is a historically constituted object—reinforced by science, institutions, practitioners, etc.—that emerged out of a particular logic, a certain political need, and historically available knowledges. And, as such, it is pertinent look at the effects of our objects of knowledge.

PART III

-

EXTENDING THE SCOPE OF ANALYSIS

Chapter 8:
The constitution and effects of safety culture as an object in the discourse of accident prevention: a Foucauldian approach

Statement of Contribution to Co-Authored Paper

This chapter includes a co-authored paper. The copyrights of this article have been transferred to Elsevier; however, authors can use their articles for scholarly, non-commercial purposes, such as the inclusion of the manuscript in a thesis or dissertation.

The bibliographic details of the co-authored paper, including all authors, are:

Henriqson, E., Schuler, B., Winsen, R. van, & Dekker, S. W. A. (2014). The constitution and effects of safety culture as an object in the discourse of accident prevention: a Foucauldian approach. *Safety Science, 11*, 465-476.

As a co-author of this paper I have taken part in the initial and ongoing discussions regarding the object and aim of this paper; I have contributed to the structure and outline of the paper; I have co-developed the theoretical and methodological approach (in particular the archaeological part); I have reviewed a (minor) part of the literature; I have contributed to the analysis and write-up of the paper (in particular the archaeological section of the paper); I have critically revised numerous versions of the paper; and I have contributed to implementing the feedback from the reviewers.

(Signed)________________ (Date) 16-02-2015
Roel van Winsen

(Countersigned)________________ (Date) 28/01/2015
Eder Henriqson (Corresponding author)

(Countersigned)________________ (Date)________
Sidney W. A. Dekker (Principal supervisor and co-author)

The constitution and effects of safety culture as an object in the discourse of accident prevention: a Foucauldian approach

Abstract

Safety culture is part of a larger discursive practice shared by scholars and practitioners in the field of accident prevention. It appears in various institutional documents, such as scientific and accident reports, industry-related statistics, manuals, and legislation. We examine the constitution of safety culture and its effects as a form of knowledge based on a Foucauldian archae-genealogical approach. Our analysis lays out the historical conditions for the emergence of the object, the scientific approaches for its delimitation, and its regular characteristics. We also describe four effects of the object in the discourse: a focused aspect of organisational culture that implies a normative homogeneity of values, beliefs, and behaviours; a disciplinary enforcement of workers' safety behaviours; a biopolitical control of organisations; and a form of governmentality that connects the conduct of individuals with organisational norms.

8.1 Introduction

Safety culture, in formal social-scientific terms, is an object of knowledge. As such, it is part of a larger discursive practice of accident prevention, together with other objects like technical failure and human error. In the last three decades, safety culture has increasingly become the focus of attention of many scientific disciplines, such as anthropology, sociology, psychology, management, and engineering, in order to explain organisational safety issues. It follows a tendency toward more 'socio-technical' and 'systemic' oriented approaches to safety. In part, this is a result of the political need to explain large organisational accidents combined with an *a priori* knowledge on organisations. Safety culture has attained an important status in the discourse of accident prevention, since it is used as explanation for accidents (e.g. Starbuck & Farjoun, 2005) and safety management performance of organisations (e.g. Hudson, 2007).

As a discursive practice, this object is part of a vocabulary shared by individuals and institutions (e.g. investigation reports, statistical information, manuals, legislations). Both recognise the need for a 'good safety culture', which often presupposes a kind of consensual agreement of values, beliefs, and behaviours in relation to risks at the workplace (Antonsen, 2009a; Edwards, Davey, & Armstrong, 2009). Individuals are called upon to be responsible for safe behaviour and, at the same time, organisations are invoked to be responsible for assuring the proper engagement with safety at the organisational level (Sibley, 2009). At the industry level, regulators define a set of practices that organisations should comply with (e.g. IAEA, 2002). The characteristics of the object are strategically taken as reference for 'measuring', 'describing' and 'comparing' the safety culture of individuals and organisations (Antonsen, 2009b; Boin & Schulman, 2008; Griffin & Neal, 2000; Lu & Tsai, 2010).

Notwithstanding controversies in the literature about conceptual definitions, methods, and measurements of safety culture (e.g. Glendon & Stanton, 2000; Guldenmund, 2000; Cox & Flin, 1998; Haukelid, 2008), it is easy to agree that as an object of knowledge it is part of a discursive practice materialised in the way institutions (including academia, industry, regulators, juridical branch, etc.) operate with it. The analysis of the object, as part of a discursive practice, "allows to de-neutralise contemporary and taken-for-granted truths and encourages a reflexive ethic-political thinking on OHS [organizational health and safety] issues" (Rasmussen, 2010, p. 476).

Despite its embryonic stage in the field of safety, the Foucauldian approach has been increasingly adopted as an analytical framework for studying discursive practices in management and organisation studies over the last decades (Barratt, 2008). Based on identified 'periods' of Foucault's work, this approach encompasses two perspectives of investigation, his so-called archaeology and genealogy (Foucault, 1980). In the former, Foucault developed a comprehensive and systematic way of examining how certain discourses gave rise to particular objects of knowledge claimed as 'truth'. In the latter, he focused on exploring the effects of power and subjectivation that particular discourses have. Foucault introduced the notion of power as a 'strategic exercise' for the control of individuals (which he called 'disciplinary power'), populations ('biopolitical power'), and as a form of control that ties individual conduct to organisational norms ('governmentality') (see Foucault, 2004). Even though organisation and management theorists have mainly focused their researches on genealogical issues of disciplinary power (Rowlinson & Carter, 2002), these perspectives can be combined into a 'Foucauldian archae-genealogical approach' in order to investigate both the constitution of particular forms of discourse and the effects of this knowledge (Knights, 2002).

In this study, we examine safety culture as an object in the discourse of accident prevention based on this Foucauldian archae-genealogical approach. The archaeological part allows a description of (a) the constitution of the object, considering the historical conditions of its emergence, the scientific approaches delimitating it, and the specifications that currently supply its characteristics. Once safety culture is characterised from an archaeological perspective, the genealogical part focuses on a description of (b) the effects that the object has as a form of knowledge, a disciplinary and biopolitical power, and a form of governmentality that connects the conduct of individuals with organisational norms. While the archaeology shows that safety culture comes from certain knowledge constructions and its rules – providing an adequate basis for the genealogical analysis – the genealogy explores some of the effects it has with interesting insights for the safety management of organisations and workers in it.

8.2 Constitution of the object from an archaeological perspective

Archaeology is way to analyse a given discourse, including the system of formation and transformation of its objects, concepts, and statements (Foucault, 2002a). It helps to describe how knowledge is produced, what characteristics it has, and what may be spoken of in the discourse considering its conditions of existence. In this study we limited the focus on the 'constitution of objects' – an important part of the archaeological project – taking safety culture as an object in the discourse of accident prevention.

Objects, such as 'human error' or 'safety culture', are artificial distinctions historically constituted and not 'scientific discoveries' of natural objects (Foucault, 2002a). Indeed, as many safety scholars have long remarked, an object like "human error" is an artificial distinction, an attribution, a human judgment – not an objective fact with immutable properties across observers or situations (Dekker, 2005; Hollnagel & Amalberti, 2001; Rasmussen, 1990b; Vicente, 1999; Woods & Cook, 2002; Woods, Dekker, Cook, Johannesen, & Sarter, 2010). The emergence of an object in a discursive practice is associated with the articulation of knowledge on politically relevant events. Accidents are a special category of events that, for multiple reasons, demand technical and social explanations (Galison, 2000; Green, 2003). Social objects, in this sense, are named and described in order to fulfil this demand.

Once named, the object receives increasing attention from scientific research when it has shown its potential for explaining a particular social order. Hollnagel (2005), for example, explains how the accident of Three Mile Island, together with the growing public concern about the risks of operating nuclear power plants, set in motion a research agenda that took 'human error' as a relevant social-scientific object. Consequently, this was followed by the development of accident models sensitive to 'human error'.

Scientific disciplines, with their apparatuses of methods and epistemologies, then proceed to delimit the object by postulating its formal definitions, its conceptual systems, and its investigation approaches. Their specification provides a singular character to the object, after which more empirical data is accumulated and the object is further corrected, modified, validated, and legitimised as part of a larger discursive practice (Foucault, 2002a). In relation to human error, for example, this practice is now so strong that the search for 'error' is a normal reaction to accidents (Hollnagel, 2005). As part of this self-reinforcing process, accident prevention strategies include a wide variety of technologies, such as error management (e.g. Helmreich, 2001) and error counting and classification (e.g. Wiegmann & Shappell, 2003).

In this study, safety culture is analysed as a symptom of a discursive practice in the discourse of accident prevention. It is a social-scientific object named, delimited, and specified according to behavioural-based approaches to safety (Tharaldsen & Haukelid, 2009). As such, it can be seen as a kind of 'discursive recodification' of other objects that are taken as 'transcendental categories', such as human error. As any other discursive object, it has been defined and investigated for analytical and practical reasons, creating a "socially authoritative interpretation of reality" (Feldman & Feldman, 2006, p. 880).

8.2.1 The emergence of the object

Objects in science do not come 'out of the blue'. The emergence of an object requires a combination of pre-existing knowledge and certain events or circumstances that demand new and appropriate explanations (Foucault, 2002a). In this sense, two conditions for the emergence of safety culture as an object are analysed: (a) the pre-existing knowledge, including a menu of concepts, methods, and disciplines related to the thematic of culture and organisations; and (b) the political need to explain important accidents that motivated the search for causes at any organisational level.

The pre-existing knowledge

Safety culture did not appear in the literature as a new scientific discovery that all of a sudden deserved to be investigated. Rather, its constitution was possible due to the existence of an important substratum of knowledge prior to the first official use of the term in the technical reports that followed in the aftermath of the Chernobyl nuclear disaster in 1986 (see INSAG, 1986, 1988, 1991). At that time, both organisational culture and safety performance were already important areas of scientific investigation.

Anthropology and sociology have studied organisational culture, at least since the 1950's, with a focus on organisational values, underlying assumptions, artefacts, rituals, norms, power, negotiations, and rites of passage. Similarly, psychological studies of organisational climate have been carried out since the 1930's (and more intensively since the 1960's). However, these psychological studies often put the emphasis on the topic of job satisfaction. In commenting on the influence of best-selling books about organisational culture in the 1980's – such as 'In search of excellence' (Peters & Watterman, 1982) and 'Corporate Cultures' (Deal & Kennedy, 1982) – Haukelid (2008) explained that these works represent mainstream ideas about organisational culture in the 1980's, when management science was declaring that, "corporation[s] with strong culture do well". These ideas, influenced by Japanese organisational philosophy, construct culture as a management tool alternative to bureaucratic control (Antonsen, 2009a). They have also contributed to seeing employees, rather than technologies, as the greatest resource of organisations and have helped to legitimise scientific work to better understand the social body of organisations.

During the 1970's, organisational knowledge was gradually integrated into the safety discourse. Based on the works of Cohen, Smith and Cohen (1975), Cohen (1977), Smith, Cohen and Cohen (1978), Zohar (1980) and others, Clarke (2000) enumerates a number of social factors to influence organisational safety performance that were already known then: top management commitment; safety as a managerial priority; status of safety officers; safety training; communication between workers and managers; environmental control and housekeeping; stable workforce; good industrial relations; and standard procedures. These factors constitute an important set of characteristics of safety culture research and managerial practice still today. In this sense, many disciplines, such as management science, social psychology, anthropology, and sociology, were providing a suite of concepts, methods, and correlated objects to explain how safety was influenced by social organisational factors (Choudhry, Fang, & Mohamed, 2007). Therefore, it is possible to assert that prior to the first use of the term (in the Chernobyl's reports) an important substratum of knowledge was available for the constitution of safety culture as an object.

The political need to explain important accidents

Accidents often show an unanticipated difference between the current beliefs about danger and the actual state of it (Mengolini & Debarberis, 2012; Pidgeon & O'Leary, 2000). In the late of 1970's and 1980's, high-visibility accidents involving risk-rich systems with reputable safety records motivated the search for explanations at an organisational level, such as the Three Mile Island (see Perrow, 1999), the Bhopal chemical accident (see Shrivastava, 1987), the capsize of MS Herald of Free Enterprise (see Sheen, 1987), the underground fire at King's Cross Station (see Fennell, 1988), the Clapham Junction rail crash (see Hidden, 1989), the Piper Alpha accident (see Paté-Cornell, 1993), and the Tenerife air crash (see Weick, 1990). Following an increasing tendency to search for causes of accidents at the blunt-end of organisations, these accidents demanded the mobilisation of organisational-related disciplines (Hale & Hovden, 1998).

As we have already mentioned, the historicity of safety culture is frequently associated with the disaster of Chernobyl in 1986 (Choudhry, Fang, & Mohamed, 2007; Cooper, 2000; Guldenmund, 2000; Yule, 2003). In the INSAG's Summary Report on the Post-Accident Review Meeting on Chernobyl Accident (see INSAG, 1986), investigators mentioned culture to explain organisational failures that contributed to the accident. From an archaeological perspective (Foucault, 2002a), it does not matter if Chernobyl was the origin of the term or if its first use was precisely associated with organisational culture research at that time. Rather, the use of the term should be considered as a symptom of a discursive practice produced by the socio-technical and organisational mainstream ideas developing at that time (Hopkins, 2006). The use of safety culture in accident reports – as are various other organisational-related terms – is evidence of a substratum of knowledge that was already being mobilised to explain safety in a strategic and tactical manner. As Brown (2004, p. 98) puts it, "reports are constructed according to the conventions of the public policy in which they are located and to which they contribute – creating, clarifying, sustaining, and modifying this particular version of reality".

Accident reports and scientific articles can be ascribed, among others, as institutionalised forms of use of the object (Foucault, 2002a). This is evident in the increased reference to the term 'safety culture' after the Chernobyl disaster. Searching for the term 'safety culture' in SCOPUS, the query returns about 2,500 references, most of them published after the year 2000. Less than 200 references appear between the years 2000 and 1991, and 10 references before 1991. The accident investigations of Überlingen (see Johnson, 2004), Continental Express flight 2574 (see Wiegmann et al., 2004), and Columbia (see CAIB, 2003; Starbuck & Farjoun, 2005) are examples of more recent uses of the term, both in scientific and technical publications. In this sense, the more the object is used in its institutionalised forms, the greater is its power as a 'regime of truth' (Foucault, 1977, 2002a). As an effect, safety culture now has an important status among objects in the discourse of accident prevention. As points Sibley (2009, p. 352), "[c]learly, safety culture has become the mantra for technologically complex and hazardous organisations".

8.2.2 Authorities of delimitation of the object

Two scientific approaches delimit safety culture as an object: (a) the interpretivist and (b) the functionalist approach. This differentiation was already made and commented upon by previous studies (e.g. Clarke, 2000; Glendon & Stanton, 2000; Guldenmund, 2000; Hauckelid, 2008; Sibley, 2009). We adopt the same approach here in order to describe how certain disciplines have been acting as authorities of delimitation of the object, thereby providing its systems of specifications.

The interpretivist approach

The interpretivist approach is characterised by scientific disciplines such as anthropology and sociology, which are often defined as the 'home' disciplines of studies on culture. Methodologies adopted by these disciplines are usually qualitatively oriented, such as ethnography and ethnomethodology, including data collection techniques based on various forms of observations, interviews, discussions, and document analysis.

This approach defines culture as something that an organisation 'does'. It considers culture as a bottom-up, complex, emergent phenomenon; greater than the sum of its parts; resistant to reductionist analysis, measurement, and engineering. For this reason, it cannot be trained and injected into individual minds. Culture is a medium by which individuals understand identity, values, beliefs, and behaviours (Haukelid, 2008). The interpretivist approach emphasises the discovery of underlying structures of meaning within an organisation (Wiegmann et al., 2004). In doing so, researchers usually use safety culture as an 'ex-post-fact' object to explain accidents, as well as the behaviour of workers and organisations (Sibley, 2009).

The interpretivist logic often defines safety as a form of expertise connected to organisational practices (Gherardi & Nicolini, 2000). In this sense, safety culture is asserted to be a consequence, rather than a cause. It is continually redefined and negotiated in relation to a larger scope of organisational factors, processes, and goals (Gherardi & Nicolini, 2000; Gherardi, Nicolini, & Odella, 1998; Sibley, 2009; Vaughan, 1999) that produce interpretation and meaning-making about danger and risk (Starbuck & Farjoun, 2005; Vaughan, 2003).

The functionalist approach

The functionalist approach is characterised by disciplines such as social psychology, management science, and engineering. These disciplines are often defined as fields of applied studies due to the pragmatism in their goals. Atak and Kingmar (2011) point out that the studies on safety culture were initially dominated by social psychology, with analyses that take safety and culture as objectively quantifiable categories. Methodologies adopted by these disciplines are usually quantitatively oriented, including surveys, various forms of questionnaires, and multivariate data analysis (Choudhry, Fang, & Mohamed, 2007; Cooper, 2000; Glendon & Stanton, 2000; Guldenmund, 2000).

This approach defines culture as something that an organisation 'has'. It considers culture a complex and multidimensional object that can be measured; a top down support for management strategies and ideologies that can be changed and engineered (Cooper & Phillips, 1995; Reason, 1997). Functionalist studies focus on the practical significance of safety culture (Choudhry, Fang, & Mohamed, 2007; Wiegmann et al., 2004). In doing so, researchers take safety culture as an object for understanding how social members signify risk and safety in order to generate senses of commitment, motivations, enhance social stability, and guide and shape behaviour (Gherardi, Nicolini, & Odella, 1998; Lu & Yang, 2011; Sibley, 2009; Turner & Tennant, 2009).

The functionalist logic assumes that there is a relation of antecedence among behaviour, beliefs, and values: values are ascribed to be at the core of a cultural system driving beliefs, which in turn are supposed to determine behaviours (Schein, 2004; Vinodkumar & Bhasi, 2010). Thus, organisations have to make workers internalise safety both as a value and a belief, in order to foment the necessary conditions for safe behaviour. This logic legitimises 'safety first' and 'zero accident' slogans and campaigns, since it is a way to disseminate values and beliefs. At the same time it perpetuates and reinforces the idea that culture is a viable characteristic, "quality that can be managed, changed or manipulated" (Haukelid, 2008, p. 95). Edwards and Jabs (2009) warn that this approach may potentially backfire, as it can prevent workers from developing a 'dissenting view', disseminating the trust on control and maintaining false believes that systems are safe.

The functionalist approach has been criticised by interpretivist researchers for invoking culture as an "iconic concept with little of the theoretical edifice sociologists and anthropologists have built" and for "instrumental and reductionist epistemologies antithetical to cultural analysis" (Sibley, 2009, p. 342). The interpretivist approach, in return, has been condemned for not offering pragmatic solutions for safety culture improvements. However, both approaches have contributed to define a set of specifications for the object, constituting the same object in different manners, with different strategies, and possibly with different effects in the discourse of accident prevention. In the next section we examine how safety culture is specified under the authority of such approaches.

8.2.3 Specifications of the object

Specifications provide the object with its singular characteristics and magnitudes that individualise it and, thereby, produce it in discursive practice (Foucault, 2002b). Even though there is an on-going debate on the specifications of safety culture as a social-scientific object, we identify important discursive regularities in relation to its definitions and its statements.

Definitions of safety culture

Many researchers agree that there is no universal accepted definition of the term safety culture (e.g. Antonsen, 2009; Clarke, 2000; Guldenmund, 2000; Cooper, 2000; Choudhry, Fang, & Mohamed, 2007). Sometimes the term is not even defined in academic studies, as mentioned by Choudhry et al. (2009, p. 996): "Only eight (8) of the twenty-seven (27) selected studies define safety culture". There is also an on-going debate on the similarities and differences between the terms safety culture and safety climate (e.g. see Deninson, 1996). Often, safety climate is derived from the basis of organisational psychology studies (e.g. Zohar, 1980), whilst safety culture is derived from a socio-anthropological basis (e.g. Haukelid, 2008). The former is frequently described as a "temporal state measure of safety culture" (Zhang et al., 2002, p. 10), while the later is often attributed to organisational factors captured by accident investigators and public prosecutors to describe accident causation (Sibley, 2009).

The term 'safety culture' is derived from organisational culture studies, where definitions of culture vary largely, including statements such as 'the way we do things around here', 'process of fabrication of meaning', 'shared learning and practices', 'shared system of meanings', 'collective programing of the mind', etc. (e.g. see Cooper, 2000, Edwards, Davey, & Armstrong, 2013; Guldenmund, 2000; Pidgeon, 1998; Sibley, 2009; Tharaldsen & Haukelid, 2009). However, a comprehensive volume of definitions summarises safety culture as 'a focused aspect of organisational culture' related to 'attitudes, values, assumptions, beliefs, behaviours' that are 'consensual, shared, and learned' by meaning-making and experience, in which a given group understand 'risks and safety' (e.g. see Antonsen, 2009; Clarke, 2000; Edwards, Davey, & Armstrong, 2013; Glendon & Stanton, 2000; Guldenmund, 2000; Richter & Koch, 2004). In a state-of-the-art research review on safety culture, Choudhry et al. (2007, p. 996) state that, "most of the definitions are relatively similar in the beliefs perspective, with each focusing, to varying degrees, on the way people think and/or behave in relation to safety." Often, definitions also imply that "organisational safety culture exists on a continuum and that organisations can have either a good or poor safety culture" (Wiegmann et al., 2004, p. 121).

It is possible to observe that safety culture is an object 'artificially detached' from a larger organisational cultural context in order to specifically address shared values, beliefs, and behaviours in relation to safety issues. This behavioural-safety oriented approach aims to promote the proper motivation, commitment, and engagement of workers and organisations with safety. Likewise, it is dedicated to understand consensual safety values, beliefs, and behaviours at the group and organisational level, rather than at the individual level. As mentioned by Schein, in an editorial note published by Hale (2004, p. 980), "if there is no consensus on key issues (…), then, *by definition*, there is no culture." Safety culture is, therefore, an object that encapsulates consensual values, beliefs, and behaviours in relation to risk and safe behaviour, but not necessarily by contrasting it with other organisational demands, such as production, quality, or cost.

Statements of safety culture

Despite the vast number of publications on safety culture over the last 20 years, there is still neither a universally accepted model nor a fixed number of variables to describe and measure it (Choudhry et al., 2007; Cooper, 2000; Guldenmund, 2000; Flin et al., 2000). Clarke (2000) and Guldenmund (2000), in reviewing safety culture studies, found important variations in the number of variables and content of them. Nevertheless, we identify three statements that currently seem to capture and address important attributes of safety culture as an object in the discourse on accident prevention: (a) safety is 'the' core organisational value; (b) safety requires management commitment; (c) safety requires workers' commitment.

According to Foucault (2002a) statements are 'functions' that display arrangements and unities of a discourse. They constitute a 'network of rules' that establish what is discursively meaningful. The 'archive' is the complex (and historical) system of statements that defines what may, and may not, be said. As such, Foucault's 'archaeology' is the study of the 'archive' (see Foucault, 2002a). Statements on safety culture thus show how safety is set out in relation to other organisational values as well as address fundamental needs in an organisational structure to better safety. Statements are taken not as assertions of scientific truth, but as a discursive mechanism that guides scientific and managerial practices in the name of safety culture. Since it puts scientific and managerial assumptions on the spot, understanding these statements is a necessary step for the analysis of the power effects that a discursive object such as safety culture has. These statements circulate in different industrial systems, such as healthcare, aviation, and nuclear power generation. The examples provided here are merely illustrative and similar evidences may be easily found in other risk-rich industries and organisations.

According to the first statement, safety culture is not just 'another' organisational value. Rather, it is an object for the analytical purpose of putting safety in the core of organisational beliefs in order to promote the "understanding that safety is the overriding priority" (Clarke, 2000, p. 76), "the guiding principle" (Wiegmann et al., 2004, p. 126). A set of institutional practices produce this statement, such as formally written, documented 'safety culture policies' and campaigns that sloganize 'safety first' and 'zero accidents'.

A symptom of this statement may be evidenced, for example in healthcare, in the guide 'Creating a Patient Safety Culture' by the Association of periOperative Registered Nurses. This document echoes frequently adopted policies in healthcare institutions when it states that, "[a] commitment to safety must be articulated at all levels of the organisation. Safety is valued as a top priority, even at the expense of productivity" (AORN, 2006, p. 940). Similarly, in aviation, airline safety policies often contain sentences like "safety is a core business and organizational value" (GAIN, 2002, p.ii). It enforces accountability at all levels as stated by the U.S. Nuclear Regulatory Commission (NRC, 2011, p. 34773): "safety culture is defined as the core values and behaviours resulting from a collective commitment by leaders and individuals to emphasize safety over competing goals to ensure protection of people and the environment". As an effect, safety is formally advertised as a moral commitment to make people across the multiple levels of organisations believe and behave according to it.

The second statement announces safety as a product of management commitment. The National Patient Safety Agency in UK, for example, states, "the level of patient safety of an organisation can be improved if there is strong leadership from the top of an organisation" (NPSA, 2004, p. 39). Similarly, IAEA (1998) advocates that "without a visible and genuine demonstration of this commitment by personal behaviour and leadership by senior managers, other workers in the organization will not be convinced of the importance of safety compared to other organisational issues" (p. 17-18). Also, in aviation, the ICAO Safety Management Manual points out:

> The ultimate responsibility for the establishment and adherence to sound safety practices rests with the directors and management of the organisation … The safety ethos of an organisation is established from the outset by the extent to which senior management accepts accountability for safe operations and for dealing with emerging safety concerns (ICAO, 2009, p. 38).

This statement is thus produced by a set of organisational practices related to continuous supervision of critical risk operations, risk analysis processes, safety training, operation monitoring, corrective actions, continual improvement of safety system, promotion of special campaigns, advertising of organisational safety values, provision of resources for safety, compliance with industrial regulations, rewarding and punishing of good and bad behaviour, coaching, participation on safety seminars and training, among others. It morally calls upon managers to engage with both the planning and the active oversight of safety activities.

The third statement asserts that safety comes from workers' commitment. Organisational practices often enforce compliance with safety working standards and operating procedures. The statement is also materialised in the use of reporting systems and workers' participation and engagement in safety related issues (e.g. committees, risk analysis teams). As emphasised, for example, in nuclear industry:

> Probably the most important indication of a good safety foundation in an organisation is the extent to which employees are actively involved in safety on a

> daily basis. If there is little involvement, with safety solely dependent on managers and safety specialists, it can be said that the organisation has failed to win people over to the safety effort. Conversely, when safety issues are identified and acted on by all employees as part of their normal working routine, the organisation can be said to have won over people's hearts and minds to the safety cause (IAEA, 2002a, p. 17).

Workers are encouraged to engage in communication about safety issues (e.g. through reporting), self-vigilance (be aware of and careful with risks), voluntarism, honesty (in relation to errors), and actively caring for each other. As indicated by Global Aviation Information Network (GAIN, 2002, p. ii):

> Each one of us will be expected to accept responsibility and accountability for our own behaviour; Each one of us will have an opportunity to participate in developing safety standards and procedures; We will openly communicate information about safety incidents and will share the lessons with others; Each of us will be concerned for the safety of others in our organisation.

This calls for 'voluntarism' to engage with safety (and the reporting of safety issues), also accompanied by concerns on how to preserve the integrity of workers so as they can speak up about safety without fear of retaliation. As pointed out by a healthcare association:

> If individuals fail to report near misses and significant events, underlying systemic issues will remain unseen and unaddressed. Without a strong and just safety culture, frontline providers and management may fail to identify an event as reportable or may hesitate to report such an event (NAHQ, 2012, p. 7).

The three statements not only encompass attributes of safety culture; but also, constitute it in its institutional forms (Foucault, 2002a, 2002b). The statements, being discursive 'functions', display an arrangement of elements (e.g. the level of management commitment to safety, workers participation and willingness to report) used for measuring, comparing and controlling safety culture. On the one hand, their attributes appear in scientific research for empirically operationalizing safety culture (e.g. Wiegmann et al., 2004); on the other hand, these attributes are part of industrial legislation to enforce safety culture into organisations (e.g. IAEA, 2002b). Both institutional forms are intertwined and reinforce each other in the constitution of the object, since the body of research is based on empirical observations guided by mainstream assumptions about social behaviour.

The relationship between the field where safety culture as a social-scientific object is developed (i.e. the research community) and the field where it is manifested, engineered, and evaluated (i.e. industries, organisations, social systems) creates a reinforcing cycle of research and application. Safety culture 'best practices' are defined in scientific reports, but at the same time these reports are taken as a legitimate and valid parameter for the establishment of industrial legislation for guiding and changing organisational practices. As a result, nowadays, it is quite usual to come across safety culture parameters that are reciprocally adopted in research and legislation throughout multiple domains – such as those found in guides for self-assessment of safety culture in nuclear installations (e.g. IAEA, 2002b) and in safety culture surveys in health care (e.g. Colla et al., 2005). It is a 'disciplinarisation of knowledge', since what we accept as valid knowledge is based on the parameters of 'truth' authorised by power relations (Foucault, 2004). This 'political economy of truth' is thus based both on scientific knowledge and on institutions that operate with it (Foucault, 1980). The relation is neither about the knowledge produced by academics while being applied in industry, nor the other way around. It is a regime of truth that crosses both institutions in line with mainstream ideas on accident prevention.

The attributes provide qualitative and quantitative parameters for scientifically describing the object. Interpretivist studies explore how and to what extent organisational artefacts, rituals, symbols, and practices reflect these statements (e.g. Atak & Kingma, 2011; Collison, 1999; Edwards & Jabs, 2009). Functionalist studies focus on quantifying the degree of which the social body of an organisation perceives safety as a value, often by means of surveys and questionnaires (e.g. Ek et al., 2007; Gonçalves Filho, Andrade, & Marinho, 2010). Industrial legislations establish standards and recommend practices that explicitly address the three statements and their attributes. Scientific and industrial practices, together, establish a regime of truth that constitutes a normative mainstream where safety culture norms (i.e. rules, standards, models) are neither completely developed by the research community nor by the organisations: they are established in the interaction between both.

The statements are part of a discursive formation that constitutes and reinforces the status of the object while circulating it in different enunciative modalities, such as accident reports, institutional documents, safety statistics, consultancy work, and scientific articles (Foucault, 1997, 2004). Based on them, organisations, managers, and employees are evaluated, certified, judged, prosecuted, forgiven, blamed, made responsible. In short, this object governs them. It is, therefore, an object with important power and knowledge effects (Foucault, 1980).

8.3 Effects of the object from a genealogical perspective

While the archaeological perspective provides an understanding about the knowledge construction of safety culture as a scientific object, including the system and rules that generate discursive regularities on it (e.g. definitions, statements), the genealogical perspective is a way to analyse the effects of such knowledge through mechanisms of power (Foucault, 1980). It does not concern itself with the search for origins, but rather refers to the study of certain forms of power and knowledge as elements connected in discursive practices. Genealogy considers the way knowledge emerges, remains, and adapts to the discourse as part of a political apparatus of power that constitutes certain conditions of existence with effects for organisations and individuals.

In this study, we describe four effects that safety culture has in the discourse of accident prevention: (a) it implies a 'normative homogeneity' of values, beliefs, and behaviours that silences conflicts, ambiguities, and differentiations; (b) as part of a set of techniques that promote disciplined safe-behaviour of workers, it has disciplinary effects that focus on the production of a useful 'docile body'; (c) it has biopolitical effects focusing on the control of populations; and (d) based on a set of institutional techniques, procedures, and statistical control of organisations, it constitutes a form of governmentality in the way it connects individual conduct to a larger biopolitical control of the population.

The first effect is derived from a 'power and knowledge' analysis of safety culture, in which, according to Foucault, the discourse allows what may (or may not) be said in a particular domain (Foucault, 1980). The second effect is based on the notion of disciplinary power (Foucault, 1977) – one of the most adopted perspectives of Foucauldian analysis in organisational studies (Knights, 2002). The third effect is based on the perspective of biopolitical power, which is focused on technologies for the control of populations (Foucault, 1997, 2008). The fourth effect draws on the concept of governmentality – which is a complex form of power that connects individual conduct with particular forms of control of populations (Foucault, 2004). The aim here is not to identify effects in an exhaustive manner, but rather to show some of the most salient (but not exclusive) symptoms of the safety culture discourse.

8.3.1 Normative homogeneity

As an object, safety culture addresses a particular aspect of an organisational system. The way it is defined implies a 'normative homogeneity' of values, beliefs, and behaviours that functions to silence important conflicts. Safety is celebrated as a moral commitment, a normative choice, that crosses all levels, but not necessarily in contrast with other organisational demands, such as production, quality, or cost savings.

Several studies have shown safety as an emergent situated practice resulting from a collective process of construction of meaning. It involves complex cultural elements, including people, technologies, and symbolic forms (e.g. Gehrardi & Nicolini, 2000; Gephart, 2004; Gehphart, Van Maanen, & Oberlechner, 2009). Safety is neither the only aim, nor even the main goal of an organisation (Hollnagel, 2009), rather, it is the condition necessary for an organisation to produce and deliver its products and services. Vaughan (1996, 1999, 2004, 2006), for example, has shown how the effects of the faster, better, and cheaper production culture at NASA contributed with normalising beliefs about risk invulnerability to the *Challenger* accident. Likewise, Starbuck and Farjoun (2005) collected several studies on the Columbia space shuttle accident, showing how safety was intertwined with larger organisational goals, practices, technologies, and routines. In any of these cases, risk taking was not solely derived by degradation of behaviour, beliefs, and values. Safety culture, as an object to explain accident causation, is part of an analytical sacrifice, which in hindsight reduces a complex social order to a limited set of artificial elements (Dekker, 2011). Several researchers challenge this kind of retrospective reductionism with different systems of explanations, using concepts such as 'fine-tuning' (Starbucks & Milliken, 1988), 'normalisation of deviance' (Vaughan, 1996, 2006), 'practical drift' (Snook, 2000), 'efficiency-thoroughness trade-offs' (Hollnagel, 2009), or 'drift into failure' (Dekker, 2011).

Prioritising safety over business and organisational decisions (Lofquist, 2010) is to ignore features of complex cultural schemes that contribute to the collective construction of safety in organisations (Vaughan, 1999; 2006). These schemes include, among others, “heterogeneity, competitive and conflicting interests, and inequalities in power and authority” (Sibley, 2009, p. 360). Edward and Jabs (2009) show that “mixed-organisational forms at work send mixed messages to workers” (p. 719). According to them, the way workers experience these contradictions may not be just an unintentional by-product of organisational culture, but rather it can produce the opposite of the intended effects – “the real importance of safety is ambiguous to managers and workers, while the role of productivity is absolutely clear” (p. 710).

Martin (1992) suggests three perspectives on organisational culture: integration, differentiation, and ambiguity. In doing so, she emphasises that organisational culture should not be seen as a solid homogeneous block of values and beliefs. Richter and Koch (2004) empirically identified evidences of these three perspectives:

> Quite clearly, it is not possible to detect a consistent unifying culture, which unite the members of the organisation on safety related questions. … [S]afety culture analysis clearly showed a differentiated set of three distinct cultures. … From the point of view of accident prevention ambiguity of intentions prevailed. … Messages from management were ambiguous, stressing in one context machining time and wastage rates, and in another, safety perspectives (p. 714-716).

Empirical studies also show that organisational dilemmas in relation to safety emerge when employees reproduce and negotiate risks and safety measures in relation to inconsistencies and contradictions in regards to production (e.g. Rasmussen, 2013).

Safety culture can be seen, therefore, as a normative homogeneity that calls for safe behaviour. It is a discontinuous aspect of an organisational system of meaning that takes the focus away from technologies and risks (Rollenhagen, 2010). It addresses the message that safety is a matter of commitment and engagement with the proper behaviour, values, and beliefs. As mentioned by Sibley (2009, p. 356), “[r]ather than a specific organisation of roles and learning processes or a measurable set of attitudes and beliefs, safety is understood as an elusive, inspirational asymptote, and more often only one of a number of competing organisational objectives". Isolating safety culture from a complex web of organisational processes silences important conflicts by setting up an agenda of moral commitment to safe behaviour.

8.3.2 Disciplinary effects

Disciplinary power can be exercised through a number of instruments: hierarchical observation, normalising judgment, and examination (Foucault, 1977). Hierarchical observation puts on display the individual's behaviour in order to improve safety at a local and general level. Several forms of observation – such as safety climate assessment or safety audits – are instigated in the name of good safety culture. Likewise, several other forms of hierarchical observation, such as local supervision, video cameras, architecture and layout of the working space, or even more sophisticated forms of monitoring that include the use of flight data and voice recorders in airplane cockpits, constitute the 'modern panopticism' of safety. Workers need to be visible, and they need to know that they are. More specifically, the later is crucial for disciplinary power – workers need to know they are visible even when nobody is watching. Because the disciplinary effect is the individual's self-vigilance, surveillance is more economical than punishment (Foucault, 1977).

Several studies have shown how performance assessment of workers is a formal managerial observation technique (e.g. Swell & Wilkinson, 1992; Willmott, 1993) that often comes with the price of resistance and conflicts (Collinson, 1998). "Tying safety to performance assessments often resulted in supervisors blaming and penalizing individuals – practices that were at odds with the espoused company policy of open communication" (Collinson, 1999, p. 586). Rasmussen (2013) demonstrates how safety culture sets to accommodate an obligation for workers to 'care for oneself' – a call that places the commitment and responsibility for risk on workers. Hierarchical observation measures workers as confessional, calculable objects, producing 'modes of subjectivation'. According to Collinson (1999, p. 580), "employees and middle managers tend to internalise the exposed values of these corporate controls". We believe it is even more than that: workers assume a mode of existence (i.e. a way of being, an 'identity') in relation to organisational safety, which is expressed in their movements, in what they say, and how they think.

Normalising judgement is supported by hierarchical observation techniques to make workers' behaviour susceptible to punishment or reward (Foucault, 1977). It is materialised in workers' normalisation of certain behaviours and beliefs through practices and habits, following sanctions from managers. Punishment and rewards are adopted as a way to approximate behaviour to a 'normalised state'. Rather than the established norms (i.e. standards, rules and legislations) for guiding organisations and workers within them, normalising judgement delineates a new social framework with which organisations and workers have to comply. It arises from the social and dynamic tension between what is and what is not considered as 'acceptable'. In this sense, safety culture has the effect of qualifying the organisation's and workers' behaviours and beliefs as responsible or irresponsible, normal or abnormal, good or bad, based on the established (moral) norm. This logic serves to highlight deviances, rewarding what is normal and punishing what is abnormal, in order to bring behaviour close to the norm (Foucault, 1977). As put by ICAO:

> Organisational culture sets the boundaries for accepted operational performance in the workplace by establishing norms and limits. … Culture sets the rules of the game, or the framework for all interpersonal interactions. It is the sum total of the way people conducts their affairs in a particular social milieu and provides a context in which things happen (2009, p. 39).

Rather than formal rules or procedures, safety culture is said to put 'boundaries' on what is, and what is not accepted; it regulates the conduct of workers by calling for 'the way we do business around here'. Values, beliefs, and behaviours are judged in relation to this artificially (but also discursively) created normative homogeneity. In scientific research, punishment and rewards are practices reinforced by being described as either an important trace of safety culture – in interpretivist studies – or as an indicator of good safety culture – in functionalist studies. In an extensive review of the literature for identifying the variables for their safety culture survey in aviation, Wiegmann et al. (2004) state that:

> One of the key components of an organisation's safety culture is the manner in which both safe and unsafe behaviour is evaluated and the consistency in which rewards or penalties are doled out according to these evaluations…. A fair

> evaluation and reward system is needed to promote safe behaviour and discourage or correct unsafe behaviour…. An organisation's safety culture, therefore, is reflected by the extent to which it possesses an established system for reinforcing safe behaviours (e.g., through monetary incentives or public praise and recognition by management and peers) as well as systems that discourage or punish unnecessary risk taking and unsafe behaviours. However, an organisation's safety culture is signified not only by the existence of such reward systems but also by the extent to which the reward systems are formally documented, consistently applied, and thoroughly explained and understood by all of its employees (p. 127).

It is not only an example of how normalising judgment is integrated in safety culture research, but also a call on organisations to document and formalise it. Edward and Jabs (2009) warn, however, that the safety of a worker is an anecdotal stance that can be "only symbolically and occasionally rewarded (…). As a result in practice, 'safe' is a necessary but insufficient characteristic of a worker whom will be punished if found wanting, but not necessarily rewarded if excellent" (p. 709). In an empirical study of safety reporting policies in North Sea oil installations, Collinson (1999) shows important symptoms of how safety incentives negatively affect reporting.

Hierarchical observation and normalising judgment are disciplinary techniques that come together in the examination, which makes the individual into a 'case', for himself and for others (Foucault, 1977). IAEA (2002b), for example, establishes the principle of self-assessment as an instrument of examination: "Self-assessment for all important activities at a nuclear plant ensures the involvement of personnel performing line functions in detecting problems concerning safety and performance and solving them"(p. 26). The reporting of safety can be seen as an act of confession circumscribed within the organisational safety culture framework of knowledge. Safety culture, as a normative homogeneity, is invoked as a reference to judge good and bad behaviour. It is the organisation's response to the incident that defines the line of what is acceptable and what is not, rather than its formal policies (Dekker, 2012b).

Safety culture has, therefore, the disciplinary effect of increasing a worker's sense of self-care and self-responsibility. Empirical studies have also shown some of these effects (e.g. Chikudate, 2009; MacEchan, 2000; Gray, 2009; Rasmussen, 2010, 2013; Zoller, 2003). As put by Tharaldsen and Hauckelid (2009, p. 376) “critical behaviours are identified and targeted for change… [P]erformance tracking and performance goals are supplemented by motivational activities and (positive) feedback systems. The right behaviour is reinforced…. The goal is to change risky behaviour into safe behaviour”. These disciplinary effects are strongly manifested in the three statements that we have discussed in the previous sections and the organisational practices encompassed by them.

8.3.3 Biopolitical effects

Where disciplinary power focuses on technologies deployed to make individual workers behave (more) safely, biopower focuses on managing a population (Foucault, 1997, 2004, 2008). It may be observed in a set of technologies adopted to compare the safety performance of organisations, based on their safety culture characteristics, promotion of industry best practices of safety management, and motivational campaigns that sloganize ‘safety first’.

Safety culture is used to compare organisations on the basis of stereotypes normatively defined according to safety culture’s ‘stages of development’. IAEA (2002a), for example, defines three stages of safety culture development: (first stage) safety based on rules and regulations; (second stage) safety as an organisational goal; (third stage) safety based on learning.

> When an organisation’s emphasis is on procedures (…) Stage 1 is likely to be the preferred choice. Stage 2 would be the choice if the emphasis were more on planning and achieving safety goals. (…) The third stage corresponds to an organisational emphasis on continuous improvement and achieving excellence. [This] evolution (…) has influenced organisations in how they view safety culture (IAEA, 2002a, p. 17).

Similarly, Hudson (2007, p. 704) proposes another model that “helps to define a pathway from less to more advanced” safety cultures based on 5 stages: pathological, reactive, calculative, proactive, and generative. These stereotypes are based on the logic that organisational safety culture can be measured based on the specifications of the object, hitherto also producing statistical parameters for industrial and managerial control. It is supposed to enable the evaluation of the extent to which the three statements of safety culture are part of an organisational life.

According to Edwards and Jabs (2009), lawsuits presume organisations to document everything possible to show a good safety culture. In this sense, a bureaucratic machinery is enforced and, in the name of good governance, training, auditing, rebuking of workers who put public at risk, and everything that could show the organisation respond quickly and generously to safety concern are documented. Workers are managed "hierarchically by means of personal testing, (...) selective recruitment, rules, and technical measures to reduce the consequences of human acts" (Rasmussen, 2010, p. 463). As an effect, safety culture is posed as part of a 'bureaucratic machinery'.

As a biopolitical instrument, safety culture impels organisations to embrace innovative safety management practices in a bureaucratic fashion (Edwards & Jabs, 2009). It contributes to preserve a bureaucratic model of organisations, encouraging them to keep working like this (Sibley, 2009). Safety management systems, said to support safety culture, are transformed into a new measure of liability, based on both the assumption of a rational functioning of organisations and the use of formal methods of control and evaluation. Organisations, in this sense, reinvent the way workers and managers are held accountable – this time for not having the proper safety culture or for not implementing safety management in accordance with legislation or the industry's best practices. "As the phenomena continually recede before efforts to control them, research advocating safety culture seems, in the end, to suggest that responsibility for the consequences of complex technologies resides in a cultural ether, everywhere or nowhere" (Sibley, 2009, p. 363).

Based on the maximisation of the body as a useful power and the minimisation of its political force, biopolitical power focuses on the control of a population (Foucault, 2008). It can be observed in the intensification of behavioural-safety campaigns, where organisations sloganize 'safety first' and try to promote 'safety awareness' as a quick-to-fix approach. It is fast and cheap to design procedures and ask workers to follow them, even though it is not necessarily the best option. Safety culture calls on organisational commitment at all levels, as evidenced in the three statements discussed in the previous section. Furthermore, as a biopolitical instrument, safety culture helps to make people believe that the bureaucratic machinery of safety works. In this sense, the greater the organisational commitment to this kind of governance, the better safety culture is supposed to be; the more safety practices and procedures an organisation has incorporated, the better its safety performance is supposed to be. Safety culture, therefore, offers a possibility for both stereotyping organisational (either poor or good) safety culture and promoting the belief that the bureaucratic model is the most viable one (Sibley, 2009). It defines a knowledge related to the government of the conduct of individuals, which Foucault (2004) called 'governmentality'.

8.3.4 Safety culture as a form of governmentality

Governmentality is a complex form of power that focuses on the government of populations by connecting the conduct of individuals to administrative standards (Foucault, 2004). It is found in the relation between biopolitical and disciplinary mechanisms, such as a calculated action over the relations that constitute the social body. As a form of governmentality, safety culture produces modes of subjectivation in which individuals understand themselves as part of an organisation. In this sense, safety culture is an expression of a safety management discourse that controls the conduct of individuals. It extends safety responsibility from the state to organisations, and from organisations to workers, so that everyone is engaged in self-responsibilization. Governmentality thus crosses all levels of an organisation. The central issue is not the coercive force exercised over those that are governed, but rather the different exercises of power that are connected in the name of good practice (Ramos do Ó, 2005).

Safety culture, when introduced at a site, relies on the notion that worker behaviour (e.g. complacency, human error) is the most frequent cause of accidents (Sibley, 2009). Such statements are often met without much opposition from the workforce, because "[p]revious discourses on human factors, errors and clumsiness have already placed the disciplined worker body at the centre of what incident causality and preventive OHS [(occupational health and safety)] work should be about" (Rasmussen, 2013, p. 92). In an empirical study centred on how senior managers seek the consent of other employees from behavioural-safety implementation, Rasmussen (2010) demonstrates how governmentality is exercised through strategies that combine reciprocal relations between biopolitics and discipline. These strategies enforce compliance with safety programmes, "managing and forming employees who are willing and able to assume responsibility for OHS programmes of self-regulation and self-care" (Rasmussen, 2010,

p. 463). According to INSAG (1988, p. 12) “an established safety culture governs the actions and interactions of all individuals and organisations engaged in activities related to nuclear power”.

During the last decades, a debate between Normal Accident Theory (NAT) and High Reliability Theory (HRT) has been exploring the controversy of the limits of safety in complex systems, such as nuclear power plants (Perrow, 1999; Sagan, 1993; Shrivastava, Sonpar, & Pazzaglia, 2009; Weick, 2004). According to NAT, complex systems require, at the same time, a centralised and decentralised form of communication and decision-making in order to be safe. This, thus, results in an unsolved paradox: how to manage the (simultaneous) centralisation and decentralisation of safety related decision-making in such organisations? HRT suggests safety culture as the element that allows complex organisations to remain safe, since it assumes that shared values and beliefs would lead to similar patterns of decisions at both a centralised and decentralised level within an organisation.

> The theoretical debate (…) is important for safety culture research because there is now some emerging evidence, (…) that highlights the critical role played by organisational culture in filling gaps in formal operational procedures in the face of complex and tightly coupled task environments (…). ‘High reliability’ culture unites commitment to mission orientation with safety culture norms, and comprises (…) a set of fundamental values; of clan, commitment to ownership of a problem by the person who finds it first, personal responsibility for activities (Pidgeon, 1998, p. 207).

Safety culture is an amalgam that connects individual decisions and behaviours into organisational standards, therefore crossing disciplinary and biopolitical domains. This connection happens in the articulation of governing techniques focused on population (such as industrial policies, best practices, and statistical data) with disciplinary instruments (hierarchical observation, normalising judgment, and examination) focused on individuals. For instance, if industrial statistics show ‘non-adherence to procedures’ as an important cause of accidents, then the disciplinary effect may be the increase of hierarchical observations to assess workers’ adherence to procedures. At the state level, the government of organisational safety management is more efficient than the controlling organisational risks; similarly, at the organisational level, the government of individuals is more efficient than penalising them for accidents.

Safety culture calls on workers to participate and, at the same time, it advocates empowerment. From a governmentality perspective, this participation is not an autonomous act of consciousness, since workers are morally convoked to take part in safety related issues. Also, they have to exercise this participation with responsibility, which in turn means 'adhering to organisational norms'. According to INSAG (1988, p. 14):

> The response of all those who strive for excellence in matters affecting nuclear safety is characterised by a questioning attitude, plus a rigorous and prudent approach, plus communication. The desired results are achieved only if the attitudes of individuals at all levels are responsive to the safety culture framework established by management.

Safety culture also presumes empowerment of workers in safety related issues. According to Wiegmann et al. (2004, p. 127) "employee empowerment means that employees have a substantial voice in safety decisions, have the leverage to initiate and achieve safety improvements, hold themselves and others accountable for their actions, and take pride in the safety record of their organisation". In this sense, there are 'agents of consciousness' inside organisations, who are supposed to provide workers with the power they need to assure safety. At the same time, the empowerment comes with moral values related to integrity and honesty, and workers are called upon to be responsible for their acts. Safety culture, therefore, builds on a logic of self-responsibilization and, at the same time, that of making others responsible.

8.4 Conclusion

The archae-genealogical approach adopted in this study provided an extensive examination of safety culture as an object in the discourse of accident prevention. In the archaeological perspective described in section 2, it was shown that, as an object, the emergence of safety culture was not due to its inherent proprieties or a positivist "reference to the ground, the foundation of things" (Foucault, 2002a, p.53). In fact, the emergence of the object is related to politically relevant events, such as the large industrial accidents during the late 1970s and 1980s, and the pre-existence of knowledge related to organisational culture at the time the object was named. Interpretivist and functionalist approaches function as authorities of delimitation of the object by defining, describing, correcting, validating, and refuting propositions in relation to it. Safety culture thus emerged from certain knowledge constructions and rules that deserve to be known. Our archaeological approach offered a way to understand the conditions of these discursive practices, including its accumulation, transformation, and discontinuities. Safety culture is not merely seen here as a consequence of a positivist and self-correcting progress of safety science, but rather as the manifestation of knowledge based on both the tensions between academic and industrial practices, fuelled by politically relevant assumptions. Nevertheless, the aim of this article was not to discuss the ontological or epistemological nature of the object, but rather to show the existence of institutional practices related to the production of it. The discursive use of the object contributes to produce its status among other objects in the safety discourse.

In the genealogical part described in section 3, we took the delineation of safety culture as a discursive object as a basis for examining how it encompasses a set of discursive practices that focus on governing workers and organisations. At the level of the discourse, the normative homogeneity of the definition of safety culture has a strategic role: it clips from a complex organisational context values and beliefs to call on workers for safe behaviour, shadowing and silencing conflicts related to ambiguity, differentiation, and all kinds of production pressures. In doing so, it takes the focus away from other factors, such as risks related to the environment and technology, addressing the message that safety is secured by a moral individual and collective commitment. In this sense, the three statements related to the specifications of safety culture, as shown in section 2, constitute the regime of truth of the object and embrace crucial components used in scientific, managerial, and industrial documents. As shown in section 3, it makes the object circulate in institutional practices, thereby reinforcing its discursive status (such as the human error of organisation in the early twenty-first century).

As an object of knowledge, safety culture has important micro-political effects of power. It produces and legitimises a bureaucratic machinery that locks the conduct of individuals into administrative standards. In this sense, it reinvents accountability in several forms: (a) increasing the workers' notion of self-responsibilization by addressing their participation and empowerment; (b) enforcing managerial supervision and adoption of innovative ‘industrial best-practices’; (c) legitimising the oversight role of industrial authorities, who will measure, compare, and regulate organisations for biopolitical control of safety.

The archae-genealogical approach presented here has important implications for research on organisational safety. First, it highlights the importance of treating safety research not dissociated of its broader organisational context that includes permanent conflicts and negotiations among several goals (safety being only one), as already advocated by several researchers (e.g. Hollnagel, 2009; Richter & Koche, 2004). Second, it emphasises the need for including the issue of power in the agenda of organisational safety studies. Other researchers have already elaborated on this, such as Antonsen (2009), who affirms that the “issue of culture and power are so intertwined that safety culture research should incorporate perspectives of power and conflict” (p. 183). We suggest that a Foucauldian approach to power – which takes it as a strategic exercise, a productive force – provides interesting insights into how certain discursive objects are constituted and display effects of discipline, biopolitics, and governmentality.

The Foucauldian approach has been adopted as an analytical framework for other cultural, managerial, and organisational studies (e.g. Alvesson & Karreman, 2000). Even thought this perspective is still in its early stages in the field of safety, further empirical studies might potentially contribute to answer questions like: how are individuals subjectivated by the safety discourse? That is, how do they assume certain modes of existence according to this discursive practice? How do the instruments of hierarchical observation, normalising judgement, and examination operate as a disciplinary power with effects on workers’ safe behaviour? How do biopolitical practices, such as statistical data, accident investigations, industry’s best practices, set an agenda of governing practices of institutions and individuals in it? And what are the effects that these governing practices have on the conduct of workers, constitution of values, and notion of accountability? Certainly, answers to these and other questions might help to explore some prominent issues on the safety research agenda, such as: the social construction of safety (see Rochlin, 1999); the normalising judgement as a mechanism of normalisation of deviance (see Vaughan, 1996); reporting as a ritual of confession, responsibilization, and participation that contributes to reconfigure the notion of self-accountability (see Dekker, 2012b); the role of surveillance systems in promoting, deflecting, and interpreting safe work practices and statistics (see Turner & Gray, 2009), to name just a few.

In order to answer these questions and further explore main issues on the agenda of safety research, future empirical studies could concretely adopt the Foucauldian approach for examining technologies, such as safety reporting, safety audits, or quality assurance programmes. These technologies are widely adopted by risk-rich organisations based on a functionalist fashion in order to provide both the visibility of dangers and the managerial control of them. Rather than just more managerial tools, these technologies offer a valuable setting for understanding safety as an evolving, political, negotiated order in organisations by displaying disciplinary, biopolitical, and governmentality effects. For example, they may highlight how safety reports, audits or quality assurance programmes contribute to hierarchical observations, normalising judgement, and examination; how managerial compliance, with non-adherence to the norms and rewards for production, dynamically delineates the framework on acceptable and unacceptable behaviours by replacing the 'thresholds of norms'; how statistical data and performance indicators govern organisations by offering a viable way for connecting industry high-level aims with organisational goals; and how these technologies reinforce particular modes in which individuals understand themselves as professionals.

Chapter 9.

On the rationale of resilience in the domain of safety

Statement of Contribution to Co-Authored Paper

This chapter includes a co-authored paper. The copyrights of this article have been transferred to Elsevier; however, authors can use their articles for scholarly, non-commercial purposes, such as the inclusion of the manuscript in a thesis or dissertation.

The bibliographic details of the co-authored paper, including all authors, are:

Bergström, J., Winsen, R. van, & Henriqson, E. (2014). On the rationale of resilience in the domain of safety: a literature review. Submitted to *Reliability engineering and systems safety*.

As a co-author of this paper I have taken part in the initial and ongoing discussions regarding the object and aim of this paper; I have contributed to the structure and outline of the paper; I have had a leading role in developing the theoretical and methodological approach; I have reviewed a part of the literature (as did the other authors); I have contributed to the analysis and write-up of the paper; I have critically revised numerous versions of the paper; and I have contributed to implementing the feedback from the reviewers.

(Signed) ____________ (Date) 16-02-2015
Roel van Winsen

(Countersigned) ____________ (Date) Jan 29 2015
Johan Bergström (Corresponding author)

(Countersigned) ____________ (Date) ________
Sidney W. A. Dekker (Principal supervisor)

Chapter 9.
On the rationale of resilience in the domain of safety

Abstract

Resilience is becoming a prevalent object in safety research and organisational practice. In this study we examine how the peer-reviewed safety science literature (a) formulates the rationale behind the study of resilience; (b) constructs resilience as a scientific object; and (c) constructs and locates the resilient subject. The results suggest that resilience engineering scholars typically motivate the need for their studies by referring to the inherent complexities of modern socio-technical systems; complexities that make these systems inherently risky. The object of resilience then becomes the capacity to adapt to such emerging risks in order to guarantee the success of the inherently risky system. In the material reviewed, the subject of resilience is typically the individual, either at the sharp end or at higher managerial levels. The individual is called-upon to adapt in the face of risk to secure the continuous performance of the system. Based on the results from how resilience has been introduced in safety sciences we raise three ethical questions for the field to address: (1) should resilience be seen as people thriving despite of, or because of, risk?; (2) should resilience theory form a basis for moral judgement?; and finally (3) how much should resilience be approached as a trait of the individual?

9.1 Introduction

Since safety science's somewhat collective conception of 'resilience engineering' during the Söderköpinge meeting almost a decade ago, 'resilience' has received an increasing amount of attention in both the academic and practical domain of safety and human factors. As such, together with other notions as 'human error' and 'safety culture', resilience (sometimes referred to as 'resilience engineering' or 'RE') is an increasingly prevalent 'object of knowledge' (Foucault, 2002a) in the scientific discourses of human factors and safety science. Leading authors on cognitive systems engineering, such as

Erik Hollnagel and David Woods, reintroduced the idea of moving away from error towards seeing both risk and safety as the product of normal organisational processes; performance variability and adaptive capacity in goal-conflicted and resource scarce environments (Hollnagel, 2008; Hollnagel & Woods, 2006; Hollnagel, Woods, & Leveson, 2006b; Westrum, 2006; Woods, 2006). As such, the resilience agenda argues for a focus on operational success and deems the study of normal work more appropriate than safety science's traditional (hegemonic) focus on failures and accidents.

Critics generally claim that the conceptual approach of resilience takes the safety field little further than already done in the late 1980s and 1990s by the school of high reliability theory (HRT) (see, for example Hopkins, 2013). This kind of criticism, which asks why we need this new vocabulary, was interestingly already pointed out in the first book on Resilience Engineering:

> What is interesting for safety is preventing accidents and not just surviving them. If resilience is used with its common meaning of survival in adversity, we do not see it to be of interest to us. If its definition is extended to cover the ability in difficult conditions to stay within the safe envelope and avoid accidents it becomes a useful term. We would, however, ask whether we do not have other terms already for that phenomenon, such as high reliability organisations, or organisations with an excellent safety culture. (Hale & Heijer, 2006)

In turn, other researchers have defended the value and novelty of resilience engineering (see, for example Ross et al., 2012). Despite these ongoing debates - whether resilience engineering merely rephrases the ideas of CSE and HRT or if it further develops these fields or even if it is a disruptive kind of innovation in safety science - due to its intuitive appeal and seemingly positive pragmatic yield, after its first explicit conception in 2006, the object immediately took off in the safety literature.

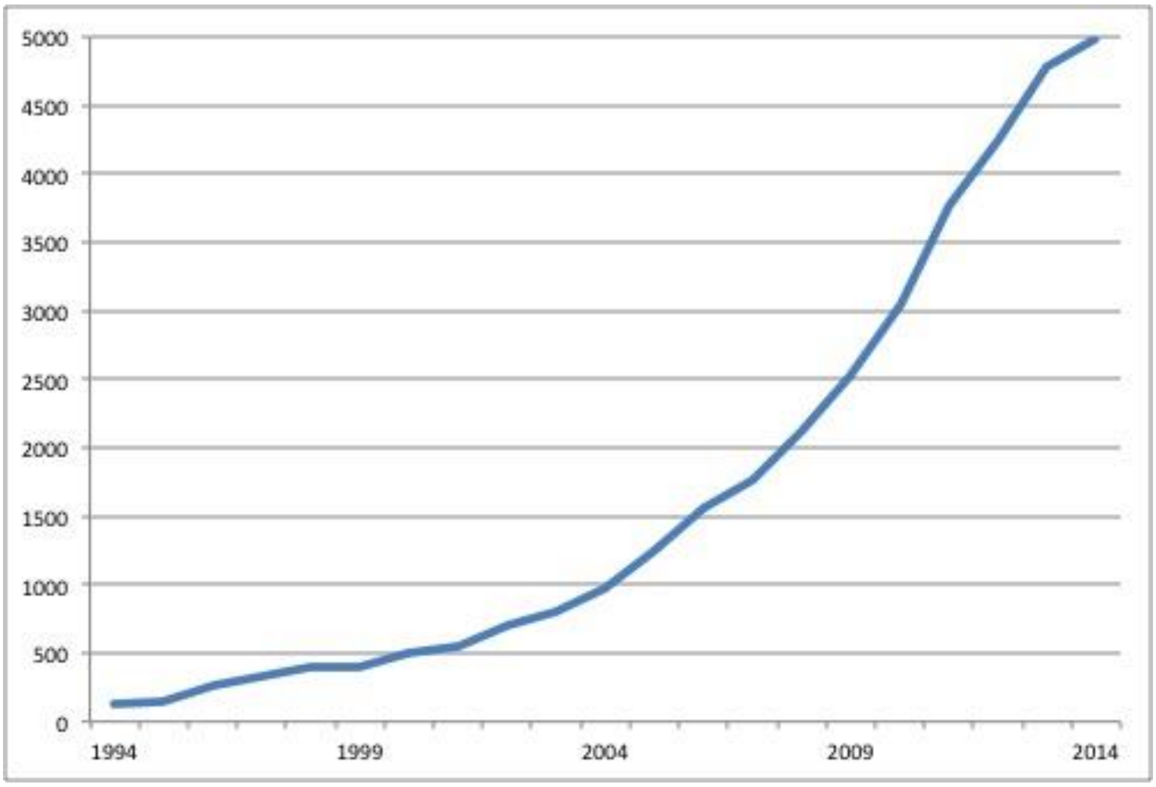

Figure 9.1. Number of papers resulting from the search "resilience" (title, keywords, abstract) in Scopus. Period: January 1 1994 to December 31 2014.

Safety science is not the only field that saw an exponential growth of resilience studies (see figure 1), as it has now become a prominent object in a number of other disciplines. The heritage of resilience as a word can be traced to Roman times to writings of Seneca the Elder, whereas its first scientific use is attributed to the 17th century writings of Sir Francis Bacon (Alexander, 2013). In contemporary academic use we relate the object of resilience to its heritage in mechanics, where it appeared in 1858 (ibid), its use in ecology (Holling, 1973; Walker, Holling, Carpenter, & Kinzig, 2004), or its use in psychology (Flach, 1988) and the health sciences (Almedom & Glandon, 2007).

The mechanical heritage of resilience can be most clearly observed in studies of resilience that adopt the stress-strain model and in studies emphasising resilience as the ability of a system (e.g. an infrastructure network (Francis & Bekera, 2014)) to regain a previous state following a disturbance. In psychology and health sciences, as a scientific object of knowledge, resilience refers to the abilities of a psychosocial subject to cope with adversity, with seminal studies focusing on the resilience of children as part of their psychosocial growth as well as their ability to cope with abrupt shocks (Alexander, 2013; Kolar, 2011; Ungar, 2005). Building on notions of systems theory (von Bertalanffy, 1950), Holling's (1973) introduction of resilience to the field of ecology in the early 1970s marked a turning point in the study of ecosystems. This turn provided the direction for a great amount of systems research, which culminated in the foundation of the *Resilience Alliance* in 2001, marking the sense of identity and community that the concept has given rise to. Influenced by the use of complexity theory in the neoliberal school of economy, the object of resilience has over the last 20 years also defined the field of social-ecological systems (Walker et al., 2004; Walker & Cooper, 2011). Definitions of resilience, in this sense, include the ability of complex systems to "absorb change without dramatically altering " (Holling, 1973, p. 7), as well as the dramatic nature of the tipping point when passing the limit of the resilient character.

Finally, the object of resilience has also emerged in other discourses such as climate change-adaptation and societal security and safety. In the latter case, following events such as the 9/11 attacks in 2001 and Hurricane Katrina in 2005, resilience found its way also into societal policy. This trend can be exemplified by campaigns such as the UN's *Making Cities Resilient*-campaign, the Australian Government's *National Strategy for Disaster Resilience,* and the *UNISDR Hyogo Framework for Action 2005-2015*.

With the widening discursive use of resilience, across all these domains, we also see counter voices and critical studies arise: is this a useful object for thinking about the reliability, elasticity or other physical properties of a mechanism, the robustness of a person or ecosystem, or even the strength of a society under adverse pressures? In fields such as societal security (Hornborg, 2013; Joseph, 2013), climate change adaptation (Schmidt, 2013), political theory (Evans & Reid, 2013) and health (Allmark, Bhanbhro, & Chrisp, 2014), we see an arising critique directed to the manner in which resilience is used and the effects that it has as a powerful object of knowledge. However, beyond the debates whether resilience is a reiteration of HRT (Hopkins, 2013), this critical stance has so far yet to emerge for the use of resilience in the safety science discourse. In this paper, we explore one possible line of reasoning for a more critical appreciation of this increasingly prevalent object in the discourses of human factors and safety science.

Objects of knowledge, such as resilience, are not 'out there' in the world, waiting for science to discover them. Instead of representing some external reality, French philosopher Michel Foucault argues that the objects of our discourses are historically contingent and arbitrary constructions; they do not mirror an external reality, but rather are the effects of certain historical discursive practices (Foucault, 2002). With his archaeological approach, Foucault aims to investigate how certain discourses—and discursive objects such as resilience—emerge and discursively function. By showing the contingent and arbitrary nature of our knowledges, as well as the effects that our discursive practices may have, Foucault aims to open up possibilities for the examination of some taken-for-granted truths.

Ten years since the Söderköpinge meeting (Hollnagel, Woods, & Leveson, 2006a), it now seems apt to assess some assumptions behind the agenda of resilience engineering. Inspired by Foucault's archaeological approach, this paper offers a study of how resilience emerges in the discourse of safety science. Based on a structured review of the literature on resilience within the safety science discourse, this paper aims to understand how resilience engineering researchers describe the rationale behind the need to be resilient (*why*), the object of resilience (*what* it is to be resilient), and the subject of resilience (*where* is resilience guaranteed, by *whom* and *how*). Eventually, we aim to initiate a critical discussion on how these three aspects combine. We will do so by raising a number of ethical questions regarding the manner in which the safety science community has so far considered the relationship between resilience and risk and the potential consequences of a normative take on resilience in combination with the 'subjectivisation' of resilience at the level of people.

9.2 Method

9.2.1 Literature review

This study was inspired by the 'systematic literature review' approach, which is characterised by its explicit research approach: the sources and search strategies for literature are made explicit and the criteria for selection and analysis of the studies are uniformly applied. This approach allowed for a transparent and reproducible strategy in the processes of articles selection and analysis.

9.2.2 Selection of literature

This study focused on the discursive use of resilience by the safety science community (typically, but not exclusively labelled as 'resilience engineering'), as opposed to the more practical uses of the object. As the acceptance of the scientific community is most convincingly guaranteed through the peer-review process (Schmoch & Schubert, 2008), we decided to limit our study to peer-reviewed academic journal articles. As an analytical choice strategy, conference proceedings and book chapters were deliberately excluded from our examination (we do realise that this is where a vast amount of the research into resilience engineering has been published). Moreover, as the number of citations for a publication is an important indicator of peer recognition (Allik, 2012; Durieux & Gevenois, 2010; Schmoch & Schubert, 2008; Sharma 2012), this study focuses on the most cited (peer-reviewed) articles concerned with resilience in the safety domain.

In systematically selecting the papers for review in our study we used Scopus. To arrive at an understanding of what outlets most resilience engineering scholars publish their work in - that is, academic peer-reviewed journals - we conducted an initial Scopus search using the following criteria: "resilience engineering" OR "organisational resilience" OR "resilience AND safety". This initial search showed that, following the Söderköping symposium in 2004 and up to December 31 2014, there are seven main journals publishing studies on resilience within the safety science community: Cognition, Technology and Work (CTW); Safety Science (SS); Reliability Engineering and Systems Safety (RESS); Theoretical Issues in Ergonomics Science (TIES); Applied Ergonomics (AE); Ergonomics (E); and Human Factors (HF). In order to identify all the studies positioned within a resilience engineering agenda we carried out additional searches on the word "resilience", either in the title, abstract, or keywords[18] for each of the seven journals. Across the seven journals we found 80 papers matching the search criteria.

The criterion to include only the papers from the main journals on resilience (i.e. the more generic journals on safety and resilience) did exclude a number of papers on 'resilience engineering' that are published in more domain specific journals. To include those highly relevant papers, we decided to lift this exclusion criterion for papers that

[18] The exact search algorithm was the following: (TITLE-ABS-KEY(resilience) AND ISSN(0003-6870) OR ISSN(1435-5558) OR ISSN(0014-0139) OR ISSN(0018-7208) OR ISSN(0022-4375) OR ISSN(0951-8320) OR SRCTITLE("Safety Science") OR ISSN(1463-922x)) AND (EXCLUDE(EXACTSRCTITLE,"Fire Safety Science"))

chiefly talk about 'resilience engineering'. As such, 6 papers were added to the data set. All 86 papers, arranged per journal, were exported into a table for systematic analysis.

9.2.3 Exclusion of literature

The analysis focused solely on the discursive use of resilience in the seventy-one selected articles. This is important, as many of the papers under analysis had 'resilience (engineering)' only as subtopic. The main topic, for example, could be the implementation of a new safety management system, and only as a secondary goal focus on how resilience could be harnessed by this system.

During the analysis it became clear that some articles were not addressing resilience and needed to be excluded. Since our study focuses on the need for resilience, the object of resilience, and the subject of resilience, we excluded papers that did not explicitly addressed these issues (papers that rather dropped the name resilience as a label without further elaborating on our points of interest). We also excluded papers that were not positioned within the field of safety science, which in the case of *RESS* implied that nine papers, focusing on critical infrastructure networks (but with no references to the core literature of safety science and resilience engineering), were excluded from the study. In total 25 papers were excluded from further analysis. In table 9.1 we outline some further details concerning the journals in which the resulting 61 papers were published.[19]

Table 9.1
Brief sampling of the papers analysed

Journal	Total number of papers matching our search criteria	Number of papers included in the study	Number of conceptual papers included in the study	Number of empirical papers included in the study
Cognition, Technology and Work	19	17	5	12
Safety Science	23	18	7	11

19 Case studies were labelled as empirical studies

Reliability Engineering and Systems Safety	21	9	2	7
Theoretical Issues in Ergonomics Science	5	5	2	3
Applied Ergonomics	4	1	0	1
Ergonomics	5	3	1	2
Human Factors	3	2	1	1
Process Safety Progress	-	2	0	2
Journal of Loss Prevention in the Process Industries	-	1	0	1
Process Safety and Environmental Protection	-	1	0	1
IEEE Engineering Management Review	-	1	1	0
Nuclear Engineering and Technology	-	1	1	0

9.2.3 Analysis and synthesis of literature

A 'good quality' (systematic) literature review starts with clear questions or categories for analysis (Petticrew, 2001; Pickering & Byrne, 2013). In this paper, the categories for analysis are inspired by Foucault's archaeological approach (Foucault, 2002) – which sees discourse as a practice that systematically forms the objects and subjects of which it speaks. As such, all the papers were methodically analysed on the basis of the following categories:

(a) the rationale for resilience: *why* do we need resilience?
(b) the object of resilience: *what* it is to be resilient, or simply, what is resilience?
(c) the subject of resilience: *who* realises or does resilience, who is responsible for resilience?

In addition to these main categories for analysis – according to which we will structure the qualitative synthesis of our analyses - we also indexed all the papers on the following variables: field of study (discipline), and type of study (method). Together with the papers' relevant bibliographic details—as provided by Scopus: author(s), title, year, journal, pages, and number of citations—the results of the queries were copied into an elaborate research matrix. This spreadsheet, with each row representing a paper and the columns constituting the categories for analysis, formed the core document for logging the analysis.

The following sections will qualitatively synthesise and discuss the three main categories of the discursive use of resilience in the academic safety literature. First we will discuss the various rationales the papers in the data set put forth, then we will discuss how the resilience is specified as an object, and finally we will look at the subject of resilience engineering studies. After the qualitative description of the 'results' of our analyses, we will turn to some ethical implications of these findings.

9.3 The need for resilience: complexity and risk

In regards to the question why the object of resilience was invoked in the papers of our analysis we identify two interconnected lines of reasoning. First, resilience seems to be an increasingly adopted, for some scholars even necessary, object to deal with the growing complexity of our socio-technical systems. Second, resilience is referred to as a manner to deal with risk, with hazards, that come with this growing complexity of our safety-critical systems. The notions of complexity and risk are sometimes connected in a seemingly (yet often implicit) 'Perrowian' manner (see Perrow, 1999).

9.3.1 Complexity

In outlining the need for resilience, the most evident observation of the papers we have analysed is that resilience gets coupled with the notion of complexity. In most of the papers we have examined, this idea of complexity is not explicitly defined or outlined, but taken as a commonsensical notion. Some papers refer to just the attributes of complexity, such as 'openness' (i.e. the lack of boundaries), 'emergence', or 'non linearity' (for formal definitions of complexity, see for example Cilliers (1998) or Page (2008)). Sheridan (2008) explicitly argues that resilience engineering, just as human error analysis, is a research focus that sprung out of a focus on complexity. As a trend, we see that some papers propose resilience as a manner of dealing with the various challenges that complexity brings.

Carmeli, Friedman, and Tishler (2013, p. 148) present resilience as a "proactive approach to safety management that recognises the complexity and ever-changing environment." Costella, Saurin, and Guimarães (2009) explain the need for a resilience engineering (RE) approach as follows: "The challenge for HS [Health and Safety] management in the context of RE is to draw up prevention strategies which adequately address complex, dynamic and unstable systems. In particular, strategies are needed which adequately take account of system variations which cannot be totally foreseen at the design stage" (p. 1057). Similarly, Brooker (2010) states that "SESAR [Single European Sky Air traffic Research Program] could be the most complex 'IT + human agent-based' safety–critical system in the world" (p. 842). Consequently, Brooker suggests that quantitative risk assessments are unable to deal with this level of complexity and he proposes to look at the resilience of the system in addition to its safety.

Andersen and Mostue (2012), in justifying the need for a resilience approach to risk assessment, draw on an established attribute of complexity, namely, the lack of boundaries. They report, that before the introduction of Integrated Operations (IO), "the boundaries of the system were easy to understand, as were the responsibility and management systems. With IO these boundaries are challenged…" (p. 2010). Tveiten, Albrechtsen, Wærø, and Wahl's (2012) argument comes from the same domain and ends up with the exact same argument, saying that distributed systems in oil and gas bring more social complexity, as well as new threats, and thus bring new challenges for emergency management.

In *CTW* several studies point to the complexities of emergency management situations, i.e. those situations in which the system is 'stretched' beyond its intended performance envelope. Nemeth, Wears, Patel, Rosen, and Cook (2011) as well as Dekker, Bergström, Amer-Wåhlin, and Cilliers (2013) use studies of non-normal situations in healthcare as the complex platform for studying resilience. Collis, Schmid, and Tobias (2013) explicitly disclose the focus of their study on the combination of complexity and non-normal conditions already in their title: *Managing incidents in a complex system: a railway case study*. Lundberg and Rankin (2013) argue the connection between complexity and unexpected conditions in emergency situations forms a need for resilient emergency management.

9.3.2 Risk

As we have just outlined, the majority of the articles under analysis refer to complexity in arguing for the need for resilience. However, complexity in itself does not justify the resilience approach. The problem that these scholars have with complexity, eventually, boils down to the idea that this increased complexity gives rise to new risks, which, in turn, demand this new resilience approach as a management strategy. This new discursive need, which arises out of the risks that come with our increasingly complex and opaque socio-technical systems, is what most of the articles under analysis - either more or less explicitly so - build their argument for resilience on.

Gomes, Woods, Carvalho, Huber, and Borges (2009) present a clear example, in which the risks of the offshore helicopter industry justify a study of system resilience. Morel, Amalberti and Chauvin (2008, 2009) do exactly the same in their two studies of fishing operations. Similarly Shirali, Mohammadfam, and Ebrahimipour (2013) state that "The increasing complexity in highly technological systems such as process industries is leading to potentially disastrous failure modes and new kind of safety issues" (p. 88). Furniss, Back, Blandford, Hildebrandt, and Broberg (2011) as well as Benn, Healey and Hollnagel (2008) locate the risk in the system environment, such as poor design, systems and processes (Furniss et al., 2011) or as a "range of risks to safety posed by close proximity to potentially hazardous processes, medicines and equipment" (Benn, Healey, & Hollnagel, 2008, p. 323) in a patient's journey through the healthcare system. Zieba Polet, Vanderhaegen, and Debernard (2010) display their joint cognitive theory-heritage in their location of risk in the complex interactions between human-machine agents. Bruyelle et al. (2014) also introduce resilience as a concept able to mitigate the risks of antagonistic threats.

There are a number of papers (especially amongst those published in Safety Science) in our study making the link between complexity and risk explicit. As we saw in the previous section, Andersen and Mostue (2012) regarded the introduction of integrated operations (IO) as capable of a vast increase in the complexity of the entire off-shore drilling system at large (including a wide range of stakeholders). Also talking about IO and the changing oil and gas industry in Norway, Skjerve, Kaarstad, Størseth, Wærø, and Grøtan (2012, p. 1952) point to 'resilient collaboration' as a new necessary manner of interacting in this complex and risky industry: "resilient collaboration, i.e., collaboration that is sufficiently robust and flexible to work efficiently and safely across the various operational states". Similarly, Tveiten et al. (2012) locate complexity and risk in the increased distribution of oil and gas systems. Steen and Aven (2011), Johnsen and Venn (2013), as well as Owen Healey, and Benn (2013) connect the need for a resilience approach to the intractable nature of risk in modern social-technological systems. Nemeth et al. (2011) operationalise this intractable nature through the notion of "gaps" in the continuity of flows when complex high-hazard systems (in their case healthcare) are operating outside of their normal conditions.

Other texts emphasise the causal link between risk and variability as a starting point for the study of resilience. This is a view that has followed the resilience engineering field since the early Ashgate volumes. Risk is seen as resulting from the same underlying processes of variability. Rather than ascribing risk as a single and distinctive outcome of variability, resilience community recognise that variability is an ever existing factor in complex systems - even in those largely standardised and proceduralised. Nevertheless, the focus in the peer-review literature seems to be on such variability as connected to the notion of risk rather than to the notion of success. Francis and Bekera (2014) state that "irrevocable uncertainty leaves risk-optimised systems vulnerable to catastrophic failures attributable to unknowable or unforeseen events" (p. 100). Similarly Wilson et al. (2009) state that "the increasing complexity in highly technological systems such as process industries is leading to potentially disastrous failure modes and new kind of safety issues" (p.787). De Carvalho (2011) as well as Re and Macchi (2010) make the same construction of resilience as the capability to compensate for risky variability.

One group of researchers has taken the resilience notion further than a manner to cope with risky complexity and variability. In both Human Factors and Safety Science, Morel, Amalberti and Chauvin (2008, 2009) go so far as to see risk as a prerequisite, a need, rather than an underlying challenge. They explicitly state that resilience and safety are far from the same thing: "Although the best safety response would be to stop fishing in borderline conditions, the resilient response is to go on, and develop survival skills, according to the situation" (p. 13). In other words, sea fishing is unsafe but done by resilient fishermen. Morel et al. (2008) locate resilience in the craftsmanship developed to cope with risk, rather than as processes to enhance safety. Moreover, as improvements in safety will eventually be balanced by production pressures, they see resilience as an effective manner of forestalling failure vis-à-vis this stand-off between safety and production. Resilience is now constructed as a capability thriving on risk. We will revisit this link between resilience and risk, and its potential ethical implications, in our discussion.

9.4 The resilience object: what is resilience

When it comes to articulating resilience as an object, scholars of resilience engineering seem coherent. In our review we make the same observation as Le Coze (2013) that the field re-conceptualises, rather than simply repeats, existing safety theories of mainly Rasmussen, Hollnagel and Woods (with Hollnagel and Woods both being previous students of 'the Rasmussian school of safety'). Few resilience scholars (re)define the scope of the concept beyond the definitions presented in the Ashgate volumes (so far there have been eight of them published). Most of the articles under review thus emphasise the challenge of resilience to cope with risky variability as the result of system complexity. More specifically, the definitions used include resilience as 'the ability to adapt to or absorb disturbing conditions' (e.g. Carmeli et al., 2013; Dekker & Pruchnicki, 2013; Francis & Bekera, 2014; Hoffman, Marx, Amin, & McDermott, 2010; Lundberg & Rankin, 2013; Nemeth, et al., 2011); resilience as 'the ability to keep the system within its functional limits' (e.g. Gomes et al., 2009; Miller & Xiao, 2007); and resilience as the four corner-stones defined by Hollnagel (e.g. Steen & Aven, 2011; Tveiten et al., 2012). Alternatively, a few papers construct resilience as the competence and know-how of people in an organisation (Andersen & Mostue, 2012; Morel et al., 2008; Re & Macchi, 2010).

Emphasising the challenge of resilience to be coping with the risky variability (also stress or disturbance) as the result of system complexity, the focus in most resilience engineering studies is on the system's adaptive capacity. This is also where the Rasmussian school of thought is introduced. The focus on adaptive capacity is rooted in a Rasmussian system dynamics model (Rasmussen, 1997), Woods and Wrethall's (2008) efforts to make an analogy between the Rasmussian model and a more mechanical stress-strain model, or Hollnagel's early (2006b) or more recent (2008) definitions that emphasise the ability to adjust system functioning prior to and following disturbance.

While most articles define resilience as a capacity to adapt to complex and risky environments, the various articles emphasise different aspects of this adaptive capacity. This section has therefore been divided into the three sub-sections that each emphasise other aspects of resilience as introduced in the reviewed literature.

9.4.1 Emphasising a link between resilience and success

Rather than ending up in a Perrowian (Perrow 1984, 1999) scepticism towards the ability to manage the risks emerging from complexity and variability, scholars are optimistically seeing resilience as the desired key to success (rather than safety) despite such risky complexity and variability. Shirali et al. (2013) offer the focus on resilience as a strategy for "how to help people to cope with complexity under pressure to obtain success" (p. 88). Costella et al. (2009) state that: "a distinctive feature of RE is its emphasis on understanding how success is obtained, how people learn and adapt themselves by creating safety in an environment which has faults, hazards, trade-offs and multiple objectives" (p. 1057). Ross et al. (2012) explain that resilience is about maintaining normal operations even during stress and disturbance. Benn, Healey & Hollnagel (2008) offer the most optimistic vision in their conclusion that a failure free (and high-quality) environment is possible by adopting a framework for engineering and controlling resilience. They explicitly refer to High Reliability Theory, which was another explicit move away from Perrowian pessimism, and show that resilience engineering aims at making teams highly reliable and consequently failure-free.

Papers that link resilience and success often emphasise the Rasmussian heritage of staying within the performance envelope, i.e. resilience as an ability to 'get the balance right' (Miller & Xiao, 2007; Wilson et al., 2009). Furniss et al. (2011) define resilience as the ability to recover and avoid accidents (the Rasmussian safety boundary) in poor circumstances. Saurin and Carim Júnior (2011) state that:

> Resilience Engineering stresses understanding how success is achieved, how people and organisations learn and adapt, and thus create safety in an environment with hazards, tradeoffs, and multiple goals (Hollnagel et al., 2006). Indeed, a key idea is that resilience is more than the ability to continue functioning when there is stress and disturbances; the ability to adjust how

people and systems function is, by far, more important from the point of view of RE (Hollnagel, 2009). (p. 355)

Gomes et al. (2009) seem to define resilience as the opposite of failing to balance risk with production pressure. Johnsen & Venn (2013) have a similar focus on the balance defining resilience "as a strategy in the risk assessment [of the key communication infrastructure used in emergency communications in railways] to improve safety, security, and quality of service" (p. 95).

As stated in the preceding chapter, Morel, Amalberti, and Chauvin (2008) have a contrasting view in their differentiation between being resilient and being safe: "the relationship between resilience and safety is much more complex than a simple, cumulative way of improving safety" (p. 3). At the same time they seem eager to appreciate the resilience strive of getting the balance right in that studies of resilience "could consider the range of controllable situations as a matter of a natural expansion of expertise and thus determine that a more resilient system is a more knowledgeable system capable of maintaining safety and gains, neither of which excludes the other, in a larger range of situations" (idem).

9.4.2 Emphasising the disturbance or stress

Earlier we concluded that several resilience engineering scholars focus in their studies on situations of crisis, disturbance, and surprise. These studies typically construct the object of resilience based on Hollnagel's definitions that emphasise resilience as the ability to adapt or absorb disturbance, disruptions and change (Collis et al., 2013; Dekker & Pruchnicki, 2013; Gomes, Borges, Huber, & Carvalho, 2014; Hoffman et al., 2010; Patterson, Woods, Cook, & Render, 2007; Skjerve et al., 2012; Steen & Aven, 2011; Tveiten et al., 2012). Such disturbance is thus central in many writings. Zieba et al. (2010) even make the link between disturbance and error, making resilience studies compatible with a focus on error. Cornelissen, Salmon, Jenkins, and Lenné (2013) highlight the link between the disturbance and performance variability: "Both resilience and performance variability recognise that adaptive capacity and flexibility to respond to unanticipated events is vital for successful performance of complex sociotechnical systems" (p. 547). Nemeth et al. (2011) distinguish resilience from control by also claiming that resilience 'happens', as an adaptive capacity, outside of the normal operating of the system.

By constructing resilience as adaptive capacity following disturbance, some resilience engineering scholars are clearly influenced by the mechanical heritage of resilience (the stress-strain model). Schraagen (2013) relies on the elaboration made by Woods (with different colleagues) in two of the Ashgate volumes (Woods & Wrethall, 2008; Woods, Schenk, & Allen, 2009) searching for analogies between the Rasmussian framework and the mechanical stress-strain theory. Carmeli, Friedman, and Tishler (2013) also emphasise resilience as a capacity for positive response and healing capabilities:

> Resilience, which is defined as 'the capacity to rebound from adversity strengthened and more resourceful' (Sutcliffe and Vogus, 2003, p. 97), is fundamental to human and organisational functioning and viability. Coping and bouncing back from experiences of failure and adversity may also be important for organisational crisis preparedness, high reliability, longevity and future growth. (p. 148)

In a similar manner, Bruyelle et al. (2014) operationalise the capability for resilience as a form of healing (absorb and bounce back through altruism) in the wake of stress.

The papers reviewed are rarely reflecting on the theoretical heritage(s) of the resilience, whether ecology, psychology, or mechanics. An exception is Francis and Bekera (2014) who spend an entire appendix on the quest of coming up with one common definition.

9.4.3. Resilience as a normative construct

Some of the papers reviewed, typically published in RESS or the more domain-specific journals, stand out with a distinct normative notion of resilience. Several of the reviewed papers (e.g. Azadeh, Salehi, Arvan, & Dolatkhah, 2014; Shirali, et al., 2013) locate resilience in indicators such as top management commitment, just and learning cultures, awareness and opacity, preparedness, and flexibility. In a similar way Paltrinieri, Øien, & Cozzani (2012) assess resilience to no less than nine lagging indicators. Johnsen and Venn (2013) adopt a normative take on the notion in their efforts to improve resilience at different organisational levels. Saurin and Carim Júnior

(2011) have a similar normative focus in their quest to develop a resilience engineering auditing system. Pasman Knegtering, and Rogers (2013) go as far in the normative take as to conclude that lack of resilience is a cause of failure: "Lack of resilience from an organisational point of view to absorb unwanted and unforeseen disturbances has in recent years been put forward as a major cause, while organisational erosive drift is shown to be responsible for complacency and degradation of safety attitude" (p. 23). Resilience, further outlined in the paper as the ability to neutralise the effects of complacency and degradation of attitudes when that leads to disturbance, deviation or erroneous acts, seems far from compatible with the purpose of the field as drawn out in the first Ashgate volume. Hollnagel and Fujita (2013) present a case study of the Fukushima disaster as a failure of resilience (mainly as inadequate anticipation of nuclear experts). Collis and Tobias (2013) also take a seemingly normative position in their study of a certain 'failure' of resilience.

9.5 The resilient subject: who does resilience

In the study of how scholars of resilience locate the subject (i.e. who or what is supposed to be resilient) of their studies we are specifically interested in the empirical investigations. Most of the articles reviewed state that resilience is something that can be seen on many organisational levels: the individual; the team; the organisation; and some take an even larger systems view and want to look at how governments influence the resilience of safety-critical organisations (the studies that are willing to take such a broad scope often build on safety/accident models that include this level of organising, such as Rasmussen's (Rasmussen, 1997) socio-technical system (STS) view). In Safety Science, several papers state this 'fractal' property of resilience (engineering): "Resilience Engineering principles may be used at any level of aggregating the cognitive system, ranging from the focus of a single worker at his workstation to the focus of the organisation as a whole." (Costella et al., 2009, p. 1057). Carmeli, Friedman, and Tishler (2013) also recognise the theoretical tendency to conceptualise resilience as a fractal concept, however, they choose to more narrowly operationalise resilience as beliefs that operate at various levels. In practice, however, we will below argue that most of the analysed studies default to looking at how individuals are able to generate these levels of resilience across various organisational levels.

In this section we will focus on the location of the resilience subject at three different levels: the sharp end, the team or management, and the functional level (actually making the connections between the different organisational levels). Our findings are summarised in table 9.2.

Table 9.2

Locations of the Resilience Subject

	Sharp-end	**Team or management**	**Functional**	**Unclear subject**
Empirical	15	15	6	8
Conceptual	4	1	5	9
Total	19	16	11	17

Note. Number of papers according to the different subject categories, empirical and conceptual papers separated.

9.5.1 Sharp-end staff - activities or skills

A third of the studies reviewed (n=19, 31%) locate the capacity to maintain system resilience at the level of the sharp-end staff: as the activities they perform or the skills that they have.

When it comes to how sharp-end staff establishes resilience through the adoption of certain strategies, the papers reviewed suggest a number of such strategies. Interesting to note is that these are typically discussed in *CTW*. Patterson et al. (2007), in the most cited study of our review, identify collaborative cross-checking amongst experts (explicitly human actors only) as a strategy to enhance system resilience. A typical focus of resilience studies in *CTW* is the human ability to be flexible as essential to the adaptive capacity of the system (Grote, Weichbrodt, Günter, Zala-Mezö, & Künzle, 2009; Owen et al., 2013) and the related focus on the tension between work as prescribed and work as performed (Norros, Liinasuo, & Savioja, 2014; Savioja, Norros, Salo, & Aaltonen, 2014). Also in CTW, Nemeth has published two articles, with different colleagues, studying resilience at the level of sharp-end of work in healthcare work. One of the studies:

> … shows how clinicians have created a consensus approach to managing their complex work domain without managerial intervention. Some approaches such as the duty call schedules are fairly formal, which others such as between-shift signouts are less so. Some are very successful, while others are less so. The varied results of these rules provide insights into the ways that clinicians manage the complexity of their work domain. It also sheds light onto the ways that operators create resilient, feasible work setting at large scale". (Nemeth et al., 2007, p. 140)

In a second study, Nemeth et al. (2011) focus on "how workers anticipate possible adverse outcomes and act in advance to avert them" (p. 199). Dekker et al. (2013) highlight the problems of attempting to manage complex situations by the use of best-practice guidelines, emphasising resilience as an alternative diversity of repertoires of the sharp-end staff. Still in the same journal, and still in the field of healthcare, Ross et al. (2012) see specialists as a key source of resilience in the system by bridging gaps, acting reactively to problems, proactively monitoring and anticipating problems, providing staff education, and patient support and education.

CTW is not the only journal publishing empirical studies that locate the subject of resilience at the sharp end. Gomes et al. (2009), in their empirical study of offshore helicopter transportation, set the stage for their study stating that: "resilience engineering, using CTA, looks at how sharp-end practitioners adapt to various types of pressures and reveals brittle points in the system" (p. 317). Cornelissen et al. (2013) focus on a specific part of the cognitive work analysis, namely the strategies used to perform work. Hoffman et al. (Hoffman et al., 2010) add to the studies of cognitive work with a focus on the way of working with a piece of technological kit as the target of resilience studies. Bakx and Nyce (2012) add to the studies focusing on sharp-end work seeing resilience as part of social accomplishment operationalised as bringing in fresh perspectives.

While the studies referred to above locate resilience at the sharp organisational end, mostly as strategies adopted to create resilience, other studies rather emphasise the skills that the sharp-end staff possess. Morel, Amalberti, and Chauvin (2008, 2009) explicitly pursue both a micro and macro ergonomic strategy for improving resilience with the subject for the interventions correspondingly being the man-machine system and the socio-technological system. However, the subject that creates resilience (or safety for that manner), is then exclusively the individual skipper in charge: "the safety of the crew and vessel depends on the fishing skipper's ability to deal with the elements, however hostile" (2008, p. 12). Re and Macchi (2010) simply refer to resilience as the operators' "competence" to deal with performance variability, whereas Andersen and Mostue (2012) locate system resilience at the sharp-end workers' risk assessments.

Collis, Schmid and Tobias (2013) present an interesting study that locates the failure to be resilient at the level of individuals at different organisational levels: "Management had failed to realise and acknowledge that…" (p.8), "both Eurostar and Eurotunnel missed signs of a developing major incident" (p.8), "the staff onboard the broken down trains … were suffering from what could be termed 'decompensation fatigue'" (p.9), "staff error had let to early loss of lighting and ventilation" (p.9), "deficiencies in central awareness of train conditions" (p.9), to provide just a few quotes. Consequently, the path towards increased resilience gets formulated as common understanding, robust plans, effective communication, and anticipating how each interfacing party would react to communications - enhanced by planning and training.

9.5.2 Management or team decision making

A number of studies analysed in this paper also locate resilience at the level of people, but at a slightly higher organisational level than the ones referred to above (n=16, 26%). Tveiten et al. (2012), Lundberg & Rankin (2013) as well as Rankin, Dahlbäck and Lundberg (2013) focus on activities to enhance the emergency management team performance, such as practicing the skills to improvise and take on additional roles in a team. Similarly Gomes et al. (2014) focus on how the emergency response team establishes resilience through (1) communication, (2) diversity, (3) small modular plans, (4) re-organising, (5) structured and integrated plan, and (6) visual supporting systems. What the team needs to avoid are sources of brittleness, including (1) no formal briefing routines, (2) not having "an appropriate workplace, (3) overload, (4) getting tired, (5) having too many team members. Furniss et al. (2011) also focus on team performance from a macro-cognitive perspective stating that: "Resilience involves interactions that happen inside and outside the head", and "so we observe interactions across tools, artefacts, people and representations ... By looking at these interactions, we focus on strategies that compensate for poor behaviour, poor design, poor systems and poor circumstances. These concerns are critical when the system has to operate outside design-basis" (p. 3).

At the higher managerial level Carmeli, Friedman and Tishler (2013) studies the top management team's ability to be proactive and adapt and Miller and Xiao (2007) argue for the need for indicators showing when management decisions take the system towards the Rasmussian boundary of functionally acceptable behaviour (risk). Also dealing with higher organisational levels, Paltrinieri, Øien and Cozzani (2012) construct resilience as an important key to "accurate risk awareness".

9.5.3 Resilience located at a functional level of the system (micro-meso-macro)

A few of the studies (n=11, 18%) address connections between different organisational levels in order to understand or improve resilience. Saurin and Carim Júnior (2011) report about the development and testing of a resilience auditing framework that addresses the individual, the team, and the organisational level. With the aim of enhancing system resilience Johnsen and Venn (2013) also suggest interventions at several organisational levels. Dekker and Pruchnicki (2013) clearly offer a macro perspective, however, their paper does not provide examples or empirical data to support or further elaborate this approach.

Schraagen (2013) offers an interesting analysis, not only in that the subject of resilience clearly is located at the macro level of the system, but also because of the selection of a research institute, rather than a high-risk organisation, as the target of analysis.

Often we note that studies may not be as functional in the manner they construct the subject of resilience as they seem at face value. De Carvalho (2011) indeed seems to offer a functional analysis of resilience, but the functions are all located at human level, with variability as a threat that needs to be limited. Similarly Costella et al. (2009) as well as Morel, Amalberti and Chauvin (2008, 2009) (discussed above) address the connection, but end up actually locating the subject of resilience at the sharp-end level. Van Westrenen (2014) offers an exception to our observation. His functionalistic approach to resilience is not only taken in the literature review, but also in the actual analysis of vessel traffic management. Similarly, Moorkamp Kramer, van Gulijk, and Ale

(2014) conduct a functionalist study, but interestingly conclude an inability of the functionalist resilience approach to address the macro nature of the system (in their case the Dutch military expeditionary organisation): “although the premises of both resilience engineering and FRAM seemed to acknowledge the dynamics of the case, they seemed to be unable to address the influence of organizational design on the ability of organizations to reduce environmental uncertainty successfully” (p. 79).

9.6 Discussion

Before we analyse the implications of how the safety sciences so far have connected the need for resilience with the way resilience is described as an object and located as subject, a couple of general remarks can be made. From our literature searches, it becomes evident that the field of resilience engineering has yet to position itself in the wider peer-reviewed scientific community. The greatest source of Scopus hits on resilience is not peer-reviewed journals, but rather the conference proceedings of the Human Factors and Ergonomics Association. Moreover, the two most influential (that is, most referenced) sources theorising resilience engineering - the Resilience Engineering Association’s conferences (there have been five of them) and the Ashgate volumes (eight so far) - are not even indexed in Scopus. Even though the resilience engineering research agenda can hardly be called new (nor heretic), research concerning the topic still does not seem to aim for peer-reviewed journals as their main outlets.

Despite the risk of various biases surrounding the studies that do end up in the peer-reviewed literature - as we have argued, in our study we have deliberately chosen to focus only on the peer-reviewed literature because it presumes a high level of acceptance by the safety science community - we see some tendencies in the publications on resilience. As also pointed out by other reviews on resilience (Haavik, 2014; Hopkins, 2013; Le Coze, 2013), it is clear that the construction of the object of resilience is done by referencing some central writings by a few authors of the Rassmusian school, mainly Hollnagel and Woods. Another interesting finding about the resilience literature is eloquently put by Alexander (2013): “it is striking how the term is used in different disciplines without any reference to how it is employed in other fields, as if there were nothing to learn or transfer from one branch of science to another” (p. 2713). As our study also shows, beyond those few popular constructions of resilience, provided chiefly by the Ashgate volumes, resilience engineering scholars seem to make little reference to operationalisations or epistemological assumptions from domains outside safety (for example holistic definitions of resilience from the ecological discourse). As such, resilience engineering has yet a lot to gain by reflecting on, exploring, importing, or just being inspired by the use of resilience in other disciplines.

Here, in the discussion of this paper, we aim to reflect on our localisations of the rational, the object, and the subject of resilience as an object in the studies we have analysed. Specifically, we will discuss the implications of the manner in which these three analytical concepts interact. We will do so by posing three questions about the discursive use and effects of resilience as an object of the safety discourse. These questions address our concern that the notion of resilience functions chiefly to load the residual risks of our complex socio-technological systems onto the backs of the individual (the front end operator or teams of operators), asking them to rely on their adaptive capacities to overcome potentially dangerous disturbances and balance safety across multiple (often conflicting) goals. We label these questions as 'ethical' because they require answers in terms of the direction scholars of resilience engineering want to take this object.

9.6.1 Ethical question 1: Resilience as thriving despite or because of risk?

As shown above, the most prevalent assumption forming the rationale for studies of resilience is the inherent risks of the complex systems of their analysis. There is a clear Perrownian heritage in such an assumption, and several of the writings reviewed in this study recognise this (see for instance Haavik's (2014) elaborate discussion on the ontological similarities between NAT, HRT and RE). Where Perrow, with his 'normal accident theory' (NAT), is concluding that, "no matter how hard we try we will still have accidents because of intrinsic characteristics of complex/coupled systems" (Perrow 1999, p. 369), resilience seems to offer a way out: adaptive capacity as a strategy to manage complexity and stay within the functional limitations of the system. Woods and Branlat (2011) state that while Perrow represents the pessimists of safety science, resilience engineering represents an optimist school of thought, one dedicated to developing "ways to control or manage a system's adaptive capacities based on empirical evidence" (p. 128). This provides another similarity between resilience engineering and high reliability theory (HRT), which also presents a more positive outlook on safety as a response to NAT's pessimism.

Resilience theory can not only be seen as an optimist approach towards the human ability to manage the risks inherent in complex systems (thrive despite risk), but just as well as an approach embracing risk as a *raison d'être* for the resilient subject (thrive because of risk). Morel, Amalberti and Chauvin (2008) provide the only study in this review that explicitly highlights such a discussion: "Should a sector's request for help in optimising production be satisfied, or should this request be denied because of the paradoxical consequences of added risk-taking, which would be the result of a successful joint assistance?" (p. 13). While resilience scholars often recognise the conflict between production pressure and demands for safe performance, this is the only study suggesting that there is an analytical and ethical choice to make. This observation is specifically interesting given that Morel, Amalberti and Chauvin's (2008) article is also the only article included in our study that, not only regards operators as accepting and adapting to the risks of the system, but also shows how the situations of "added risk-taking" are the ones where the resilient subject thrives. Not included in our literature review (as it is published into the field of political science rather than safety science), but highly relevant for our discussion here, Evans and Reid (2013) conclude in the new journal *Resilience:*

> Life quite literally becomes a series of dangerous events. Its [the resilient subject's] biography becomes a story of non-linear reactions to dangers that continually defy any attempt on its behalf to impress time with purpose and meaning. As the resilient subject navigates its ways across the complex, unknowable and forever dangerous landscapes that define the topos of contemporary politics, so the dangerousness of life becomes its condition of possibility rather than its threat. In a certain sense, the resilient subject thrives on danger. (p. 87)

The ethical question at hand asks the research community to elaborate on the relationship between resilience, risk, and safety. The answer will have implications for the level of risk-taking that is seen as a necessary prerequisite for the ability to prove resilient in complex socio-technical environments. Recognising the idea that risk is inherent in system complexity, whether drawing pessimist or optimist conclusions, is also recognising that the "zero harm" or "zero accident" is a fallacy and that the definition of safety should always be considered in relation to an acceptable level of risk.

In this case, whether the acceptable level of risk is pre-defined or not, the subjects of resilience are supposed to consent to it, and, at the same time, keep it at the appropriate level while negotiating multiple goals (e.g. production, protection, quality, workload). This does not mean denying the ontological nature of risk in complex socio-technical systems, but rather it means to recognise the discursive power of safety in its innovative categories (i.e. resilience) that helps society justify and accept political choices for operating certain technologies, under certain circumstances.

9.6.2 Ethical question 2: Resilience as a basis for moral judgement?

The early conceptual writings on resilience engineering (in the Ashgate volumes) stress that the resilience perspective implies seeing safety and risk as emerging from the same processes of performance variability. This is used as an argument against focusing on negatives, such as error or poor judgement, as explanations for organisational failure. With this in mind it is somewhat surprising that one of our observations reviewing the literature is how several resilience theorists approach the notion normatively; i.e. resilience as a characteristic that a system needs to possess in order to stay within functional limits and that there are indeed qualitative differences between the adaptive processes of variability that lead to success and those that lead to failure. The question then becomes whether such a normative view can be sustained without also retreating to descriptions of negatives – 'failures' - to maintain system resilience as causal explanations for accidents. In this case, resilience seems to become a stand-in for the notion of safety.

There is a risk that also resilience becomes another normative reference for moral judgement in the wake of organisational failure - just as notions as safety culture, human error, organisational sensemaking, and many other object of the safety discourse. When Hollnagel and Fujita (2013) construct the Fukushima disaster as the result of a lack of resilience (including descriptions of overconfidence, complacency and forgetting to be mindful), they (uncharacteristically for Hollnagel) describe the accident in terms of negatives. The important ethical implication of this kind of construction, however, is how the accident becomes the consequence of not following resilience principles (i.e. the four corner stones). As such, resilience can become another *ex post facto* manner to tell the moral tale of how those who did not follow the principles (of resilience) caused the accident - just as Weick did with the twelve fire-fighters who perished in the Mann

Gulch fire when they did not follow the principles of 'organisational sensemaking' (Weick, 1993), or as the multiple accident investigations blaming the failed system state on the (moral) failure of the pilots to maintain 'situational awareness'. The issue here is not ontological or epistemological, but philosophical: social-scientific categories such as resilience are historically contingent constructions that have certain discursive effects. They are not natural, nor are they the result of scientific progress (Foucault, 2002). As such, these categories are not neutral as they imply a particular understanding that is grounded in our moral assumptions.

9.6.3 Ethical question 3: How much is resilience an individual's trait?

The final ethical question that we would like to pose is tightly connected to the two asked above. If resilience theory, at the same time, (1) embraces an optimistic view of human adaptive capacity to keep the complex (and inherently risky) system within its functional limits, and (2) holds the possibility to construct accidents as a moral failure to stay within such limits, we need to ask who is the subject that makes this call. Even though theoretically resilience is typically conceptualised at the functional level of the system, our conclusion from reviewing the peer-reviewed literature is that the majority of the empirical studies of resilience, within the safety discourse, locates the subject of resilience at the level of individuals (sharp-end staff or management decision makers), rather than the system. In our data we can see that whereas the conceptual writings to a greater extent locate the subject of resilience in the relationships between humans at different organisational levels and the resources that they possess (such as routines and technical artefacts), the empirical papers tend to locate the subject of resilience at the level of the individual (human) actor *despite* their configuration in their messy systems (their routines and limited resources). That said, this is to the extent that the papers at all offers a clear location of the subject of resilience. As is shown in table two a vast amount of, especially the conceptual papers, are not explicit at all regarding the location of the subject of resilience.

This discussion highlights several difficult balancing acts and analytical (ethical) choices to be made in resilience studies. It is not easy to stay true to the premise that there are no qualitative differences in the organisational processes of performance variability that emerge in either success or failure. It is not easy to construct resilience at the functional level of the system. In combination the three ethical questions raised ask for a wider discussion concerning the implications and effects (intended or not) that resilience is having as an object of the safety discourse. Even an optimistic approach towards the human ability to adapt to emerging risks must include a discussion of what risks we should be more pessimistic towards and not expose our (individual) operators to. Even a community dedicated to studies of adaptive processes resulting in success must include a critical discussion of how to (ex post facto) treat failure and the people involved in the organisation that failed.

9.7 Conclusion

Our conclusions are based on a literature review studying how resilience has been constructed as an object of knowledge in the safety sciences. The object of resilience seems to offer an optimistic perspective in regards to the (sharp-end) operators' abilities to guarantee the safety and success of modern socio-technical systems through their individual abilities to adapt to the inherent risks of the environments in which they are placed. We propose the need for a critical, yet constructive debate, reflecting on the underlying assumptions and epistemology of resilience (for example by looking at how other scientific domains operationalize this object), and, perhaps more importantly, the implications that come with its discursive use. We are not asking for a repetition of the NAT/HRO-debate, even though this provided a more nuanced view of the field during the 1980s and 1990s, as we are not advocating a pessimist conclusion regarding the abilities to manage complex and coupled systems. Rather, we propose a discussion on the discursive and ethical implications concerning the role that this object creates for the subject to fulfil.

Conclusion

This thesis started with the case of Karl. In particular with my initial bewilderment and later frustration over the discursive use of 'situation awareness' in the Surpreme Court's verdict (2013): "Maintaining situation awareness is key to proper navigation". I wondered, why is this sentence included in the verdict of a criminal case? What purpose does the normative use of this object serve here? Given the aims of human factors and safety science (see chapter 1), was situation awareness (SA) not conceived as an object to increase the efficiency of our socio-technological systems from a systems perspective? Moreover, in this case I fail to see how the object—or the court case for that matter[20]—contributes to optimise human wellbeing.

To avoid criminal liability for negligence professionals must act in line with accepted—normal—practice in their field of practice (see my discussion of the 'Bolam test' in chapter 2). This is where human factors and safety science come in. Some of the discursive objects that they promote, such as SA, are taken to establish exactly these normative benchmarks for behaviour. Take, for example, the following sentence that opens a recent article in the journal of Anaesthesiology: "Accurate situation awareness (SA) of medical staff is integral for providing optimal performance during the treatment of patients" (Schulz et al., 2013, p. 729). As dictated by this leading journal in the field of anaesthesiology—which thus has a significant amount of power to set standards—the 'ordinary skilled anaesthesiologist', as per the Bolam test, is now required to have 'accurate SA' (at all times) in order to avoid liability for negligence.

The same effect features in Karl's verdict, in which the argument for Karl's dereliction of his duty pivoted on the expert's standard—that of the 'reasonable maritime officer'—that dictated that, "[m]aintaining situational awareness at all times and in all circumstances is key to proper navigation" (Supreme Court of British Columbia, 2013). Legitimised by human factors and safety science publications from influential institutions or governmental bodies, maintaining accurate SA is now part of the

[20] Similarly, I fail to see how the criminalisation of Karl serves the purposes of criminal justice—whether retribution, rehabilitation, prevention, or deterrence.

‘ordinary skills’ of the ordinary maritime captain, thus making them legally liable for losing SA. This discursive use of SA, as setting a norm for which people can be held criminally liable, seems to work counter to overall intentions and purposes of human factors and safety science.

This is where my project started. In feeling the need for some critical reflection on the discursive practices that the disciplines of human factors and safety science promote—in this case, the object of situation awareness—I turned to Michel Foucault. In his archaeological projects, Foucault (2002a, 2002b) hopes to capture the historically contingent ‘conditions of possibility’ and ‘rules of formation’ that organize the production of discourses and so delimit what can be said and thought at a particular time and place. In doing so, he aims to show that our knowledges are not the necessary culmination of (scientific or societal) progress, and that things could thus be different.

Amazed by the lack of reference to this influential author, this thesis project pursued the dual aim of introducing Foucault’s archaeological thinking to the disciplines of human factors and safety science, whilst, by doing this, providing reflections on some of the discursive objects that these disciplines promote—specifically the objects of situation awareness, safety culture, and resilience. This thesis asked what insights Foucault’s archaeological analyses of objects can provide these disciplines, and if these can be used to reflect on the some of the discursive practices these disciplines are promoting. Here, in the conclusion, I aim to discuss how the thesis, as a whole, answers these questions. The conclusion also aims to ‘translate’ some of the rather Foucauldian arguments to the audiences of the disciplines of human factors and safety science.

First, I will briefly discuss how we can look at the ‘insights’ that have resulted from bringing Foucault’s archaeological approach to the disciplines of human factors and safety science. Hereafter, I will turn to the question of how these archaeological insights, as ‘critiques’, can help to critically reflect on some of the practices that these disciplines are promoting. These theoretical sections of reflection will be followed by the practical and ethical question of why any of this should matter to the (applied) disciplines of human factors and safety science. In taking these steps, I hope to more convincingly bridge the vast—epistemological and practical—gap between Foucault’s archaeology and the applied disciplines of human factors and safety science. Finally, I will briefly outline some of the limitations of my PhD project, as well as suggest some directions for future research.

1. Archaeological 'insights'

In asking what insights Foucault's archaeological analysis of the formation of objects can provide the disciplines of human factors and safety science, the question implicitly also asks *how* we can use Foucault's archaeological approach to reflect on some of the discursive practices that these disciplines promote. As discussed at length in Part II of this thesis, Foucault does not provide an instrumental, or prescriptive, method for 'doing' archaeological analysis. Rather, through his various archaeological books and papers on discourse, he presents a number of complex, intertwined, ideas and suggestions for archaeological analysis—or better, analysis of discourse at an archaeological level of description. Chapters 3 and 4 have tried to tease out some assumptions and concepts related to the concept of discourse and Foucault's archaeological approach. From these elaborate descriptions, I have focused on a number of 'analytical concepts' for the archaeological analysis of objects, primarily those of the 'rules of formation' and 'conditions of possibility'.

The four analytical studies in this thesis, presented in Part III and IV, draw on a number of these analytical concepts for analysis. As such, these studies present instances of how Foucault's archaeology can inform these more specific studies. It is, however, important to state that these specific studies present merely possibilities of analysis; the analysis could have gone in a million different directions. Not only could the studies have drawn on different analytical concepts—a study could have focused more explicitly on the 'systems of dispersion' rather than the 'rules of formation'—but also the specific manner in which the analytical concept of choice was pursued is not set in stone. For example, instead of describing the three conditions of possibility for the emergence of SA, see chapter 7—the logical of representation, the political need to focus on the social contributions to accidents, and the cognitive knowledges that serve to as a substratum of knowledge from which SA could emerge—I could probably have located other, or more, discursive regularities as conditions of possibility in the scientific literature in which SA surfaces.

Furthermore, the studies could have selected other materialities to study the discursive formation of SA, safety culture, or resilience. They could have included or focused the analysis on institutional or governmental documents that constitute this object, rather than the scientific literature—in particular peer-reviewed journal articles. Only the study into safety culture, chapter 8, includes the analysis of some institutional documents. The study of the conditions of possibility, chapter 7, does so only to a minimum extent. Alternatively, given that most of the 'state of the art' research on resilience is published in the (eight) Ashgate volumes, these books could have made for a convincing materiality to study this object. All of these choices would have yielded different archaeological 'insights'.

Even though (my co-authors and) I had justifiable reasons for the analytical choices that were made in the various studies[21], these 'reasons' are only 'justifiable' in light of the current norms and standards for (Western) scientific practice. As such, the insights that the analytical studies present are just as historically contingent and arbitrary as the knowledges—discursive objects—that they analysed. This is why Foucault labelled his work as 'fiction' (see chapter 5). It is therefore important to explicate that the archaeological studies in this thesis have not 'discovered' some deeper or true insight, but rather they have constituted a number of conditional—one could even say provisional—'insights'. Not only are these 'insights' contingent upon the current norms for scientific research, the materialities selected for analysis, or other analytical and methodological choices, but they are also dependent on a host of cultural and personal preferences, the requirements for journal publication, and other local conditions and resources. In other words, the specifics and the context of the studies are important, and, consequently, the 'insights' should not be considered without taking these into account. As such, each archaeological study presents its own unique 'insight'.

To stress the point, the 'insights' for critical reflection that this thesis is looking for, do not present any (representative) insights into the object per se, let alone into the discursive practices of human factors and safety science at large; the studies do not provide some true or metaphysical knowledge about the respective discursive objects, but simply provide archaeologically inspired and context-dependent[22] descriptions of

[21] Admittedly, these analytical choices are more clearly and explicitly addressed in the two more systematic studies, chapters 6 and 9.

[22] In the largest possible understanding of the word 'context'.

the ‘discursive regularities’ that these studies have found in the specific materialities in which the object emerges. Interestingly, the significant role that context plays in the final form that the archaeological studies take, matches human factors and safety science’s core principle of systems thinking. As Moray explains in his appropriately titled manuscript, ‘“De Maximis Non Curat Lex” or how Context Reduces Science to Art in the Practice of Human Factors’: “Unlike many other engineering disciplines human factors is extremely sensitive to context. The result is that there are no such things as context free laws in applied psychology…”(Moray, 1994, p. 526). As such, disposing of the generalisable ‘truth-value’ of the ‘findings’ of my archaeological studies, in this manner, should thus not be too difficult for the disciplines of human factors and safety science to accept.

In short, there is not one—best—way of rendering archaeological analysis into a more specific research agenda (for human factors and safety science). Nor are there true insights to be *found*. Not only does archaeological analysis not intend to provide a definitive or true account, but with Foucault’s archaeological conception of discourse, there is simply no deeper truth or meaning to ‘see into’—‘insight’, in this sense, is a profoundly un-Foucauldian word. As such, the conditional insights presented by the individual studies cannot be reduced to their essential points or key insights. The part of my overarching thesis question that asks about the insights that Foucault’s archaeological approach to study objects can bring to the disciplines of human factors and safety science is thus answered in the various archaeological studies themselves—as these should be considered in their ‘own complexity’ (Foucault, 2002a, p. 52).

The contingency or context dependency of the archaeological ‘insights’ that this thesis provides, however, does not mean that they are not useful (see the next section). It simply means that from an archaeological perspective, the insights that the four studies have presented are not to be assessed on the basis of their ‘truth-value’. How then, should we evaluate whether this thesis’ goal of bringing archaeology to the disciplines of human factors and safety science is a worthwhile endeavour? One possible criterion for such ‘meta-reflections’ is to ask, in paraphrasing Lévi-Strauss (1966), “is this good to think with?” Are these ‘insights’ good to think with? To specify slightly, do the various archaeological studies provide new, interesting, or different ‘insights’? Do they provide alternative ways of thinking about these objects? These questions, however, are not up to me to conclusively answer—that would be replacing one truth win another. It is merely my hope that the reader, as he or she has gone through these papers, has somehow or somewhat expanded his or her thinking on these objects that feature quite dominantly in the domains of human factors and safety science.

It is also my hope that not only these specific 'insights' but Foucault's archaeological approach in general will help to contribute to a plurality and multidisciplinarity—both theoretically as methodologically—in human factors and safety science. As explained in the introduction, one of my chief aims for this thesis—in bringing the (archaeological) work of Foucault to the disciplines of human factors and safety science—was to add to the plurality of thinking, to the 'epistemological pluralism', in these discipline disciplines. In line with my reflections of the contingency of the studies' 'insights', Healy (2003, p. 594) explains:

> 'Epistemological pluralism' surmounts the constraints imposed by adherence to narrow representational perspectives, and the methods that attach to them, by legitimating and facilitating the deployment of other relevant perspectives and methods in parallel with them. This liberation centres upon a recognition of the fundamental significance of the context of the production, dissemination and application of knowledge by focusing upon the critical role of practices, methods and their institutional manifestation. This should not be interpreted as a denial of the veracity of scientific insights but is rather to highlight their partiality and how, consequently, they are best deployed in contexts not structured and determined by science and that allow other knowledge to complement it.

The bottom line for human factors and safety science is (typically) said to be their applicable recommendations—"safety science is an applied subject whose ultimate justification must be that it makes things in the field safer and healthier" (Hale, 2014, p. 68). Wilson (2000) even explains that some ergonomics see application as its very *raison d'être* and that the term 'applied ergonomics' is thus seen as a tautology. As such, ever since their first conception, human factors and safety science typically draw on the positivist 'tools' of engineering and experimental psychology, aimed at generating 'true' and 'definitive' knowledge that is directly applicable.

However, as briefly touched upon in chapter 1, this epistemological insistence on generating (causal) insights and true universal knowledge may not suit the increasingly complex and post-modern problems that human factors and safety science are currently facing. In our modern, rapidly changing, complex and globalised world, this (single) approach might no longer be enough: "The interactive complex of modern life provides the setting for ergonomics and we can no longer understand human behaviour and performance as we once did" (Wilson, 2000, p. 563). Batteau (2002) provides some explicit examples (in regards to aviation safety), and explains that the very approaches that have brought the discipline of human factors success in the past, may now start having harmful effects: "These strengths now leave the paradigm vulnerable to an array of over-the-horizon challenges. These challenges include Third World aviation, non-western conceptions of management, the changing construction of the flying public, and a metastasis of organisational complexity within commercial aviation" (Batteau, 2002, p. 148).

In dealing with these 'over-the-horizon' challenges, other—perhaps less practically oriented, but also less linear, more reflective, and epistemologically diverse—approaches are needed to enrich the human factors and safety science 'tool boxes'. Batteau forcefully argues that socially oriented forms of inquiry that "rarely achieve finality", such as anthropology, "are no less critical to flight safety than more positivist inquiries into fatique or situational awareness" (2002, p. 148). Similarly, as briefly explained in chapter 1, Diane Vaughan (1996) provided a sociological explanation of the Challenger accident, which lead to new ideas—such as the 'normalisation of deviance'—that were not only 'good to think with' but also resulted in new practical suggestions for (understanding) human factors and safety science. Seeing—describing—things in new ways, based on new theoretical frameworks, and new epistemologies may, very well, eventually lead to new practical knowledge.

As such, I hope that this thesis provides a small step in showing that Foucauldian (archaeological) analyses can also create 'insights' and supply new ways of thinking that can contribute to the epistemological pluralism of human factors and safety science. "Recognition of the ways in which knowledge acts to shape our world, culture, institutions and actions, and embroil us in complexes of both people and things paints a complexity remote from the certainties conveyed by representational understandings" (Healy, 2003, p. 693). Modern science's 'will to truth' (Foucault, 1981)—its (narrow) epistemological insistence on aiming for true and definitive accounts and the limited set of associated methods that accompany it—not only results in narrow understandings of the problems that we are currently facing, but also silences and represses other ways of thinking. It is this kind of (epistemological) dominance that Foucault's approaches may be able to dispel. It is my hope that by bringing Foucault's archaeological approach to the disciplines of human factors and safety science, by legitimising and facilitating 'other ways of knowing' (Healy, 2003), the thesis helps to counter the constraining and counter-productive effects that come with the hegemony of these disciplines' positivist approaches.

2. Archaeological reflections on human factors and safety science

The second part of the thesis question asks if we can use the archaeological insights—the unique and contingent 'insights' that are presented in the individual studies—to critically reflect on some of the discursive practices promoted by human factors and safety science. In order to address this question, I first need to specify what I think such 'critical reflection' entails.

As outlined in the introduction, 'a critique', for Foucault, "does not consist in saying that things aren't good the way they are. It consists in seeing on what type of assumptions, of familiar notions, of established, unexamined ways of thinking accepted practices are based" (Foucault, Michel 2001, p. 456). In a way, this is exactly the kind of knowledge—'insights'—that the various archaeological studies have provided. The four studies provide 'critiques' in that they outline, at an archaeological level, the assumptions and familiar notions on which the objects of situation awareness, safety culture, and resilience are based. In order to use these critiques to critically reflect on the discursive practices that human factors and safety science are promoting, I ask whether these objects function, at the archaeological level, in line with the explicit goals and assumption of the disciplines of human factors and safety science as a whole (as outlined in chapter 1).

Even though the four analytical studies each provide their own contingent 'insights' on the objects that they study, in order to critically reflect on the disciplines of human factors and safety science as a whole, it may be interesting to have a look at what they have in common. Note, however, that in no way this should be seen as a form of 'triangulation' in order to establish some underlying true insight or critique. Rather, I suggest to see these 'critiques' as possible narratives—thus disregarding questions about their truth.

Having said this, the four analytical studies all conclude that the respective object of their study is historically constituted—that is, they emerge under certain contingent conditions, rules, and relations—and as such, that they have certain effects (see also next section)[23]. Another aspect, or 'insight', that these studies have in common as critiques is that they all seem to construct the human individual as the central locus for intervention in our complex socio-technological systems.

According to both SA studies, presented in chapters 6 and 7, the object of SA emerges in the literature to better design the system or interface for human-machine interactions in complex and highly-automated socio-technological systems. However, from an archaeological perspective, SA is not just the result of this discursive need to address the social part of our socio-technological systems, but also constitutes and reinforces it: SA makes the human individual, in particular his or her 'awareness' of the system, more susceptible to socio-scientific analysis.

The object of safety culture seems to discursively function in a similar manner. Even though explicitly safety culture seems dedicated to understand (consensual) safety values, beliefs, and behaviours at the group or organisational level, its behavioural-safety oriented approach aims to promote the proper motivation, commitment, and engagement of workers at the individual level. By tying individual conduct to organisational norms, the object thus functions to turn the focus of safety away from more systemic factors, such as risks related to the environment and technology, towards reinforcing the idea that safety is secured by the moral commitment of (a collections of) individuals.

The study into the object and associated subject of resilience provides a similar characteristic about the object's discursive functioning. It argues that resilience is constructed, in the safety literature, as people's capacity to adapt to emerging risks in inherently complex and dangerous systems. As such, the subject of resilience is typically the individual, either at the sharp end or at higher managerial levels. Even though

[23] However, as this is a main assumption for archaeological analysis, this is not a surprising 'insight'—if it can be labelled as an outcome of the studies at all.

resilience is often explicitly conceptualised at the functional level of the system, as a discursive object it functions to place the commitment and responsibility for risk on individuals.

In short, the archaeological analyses of the objects of situation awareness, safety culture, and resilience, argue that these objects discursively turn the focus onto the human individual as the locus for intervention for dealing with the various human factors and safety challenges in our socio-technological systems. This (shared) archaeological critique thus seems to undermine the explicit intentions, in term of systems thinking, that human factors and safety science have with these objects.

The fields of human factors and safety science were built on the work of pioneers as Chapanis and Fitts and Jones, who claimed that the features of people's equipment and environment shape human performance, and, as such, the interventions (and discursive practices) of these disciplines should not be directed at individuals but rather at the level of systems (see chapter 1). As such, the discursive practices that human factors and safety science promote explicitly aim to move away from a focus on the individual towards a systems approach. In this light, the object of SA, for example, claims to help with the design of human-machined interfaces of highly automated systems. Similarly, the objects of safety culture and resilience, explicitly aim to improve the safety at the organisational—or systems—level. However, in critically reflecting on the discursive practices that human factors and safety science promote—here in the form of these three prominent discursive objects—this is not what the archaeological studies 'find', or better 'construct'. According to the studies in my thesis, the major focus of human factors and safety science for intervening in our increasingly complex socio-technological systems is the (self-contained) individual human subject, rather than the system.

One possible explanation for this focus on the human individual in our socio-technological systems may lie in the inherent contradiction between human factors and safety science' pragmatic orientation and the systems approach they try to build upon. As a practical discipline, human factors aims to provide (design-) recommendations to improve the interaction of humans and systems. Safety science, similarly, hopes to contribute to making our world safer, in whatever practical manner. However, while (reductionist) science often isolates its objects from their context—to study them in controllable experimental settings—in order to be able to provide causal explanations, systems approaches rely on a more intuitive and holistic approach that does not yield the same 'practical' truths (in terms of cause and effect). Systems explanations point away from single-cause explanations, and seek multiple diverging narratives that enlarge the scope of analysis to understand emergent phenomena. As small changes may have large effects in systems theory, at best, it hopes to find coexistence and correlation—as opposed to causation—in the interactions and relations that make up our socio-technological systems. Consequently, after fixing the obvious, manifest problems in the system, it becomes increasingly difficult for these disciplines as systems sciences to provide 'simple' and practical solutions: where do we intervene or re-design systems when everything in the system is interconnected and interdependent? When we cannot locate a cause for the accident, how can we fix the system so that things do not go wrong again? Perhaps, unintentionally and unknowingly, objects as situation awareness, safety culture, and resilience, despite their explicit reference to systems thinking, have been unable to break free from this practical need to find a locus of cause, or at least a locus of intervention.

In summary, even though human factors and safety science explicitly aim to rely on a systems approach (see chapter 1), some of the discursive practices that they promote (themselves!) may take the individual (operator) as the central locus for intervention. These are the "unexamined ways of thinking accepted practices are based" (Foucault, Michel 2001, p. 456) that archaeological analysis can help to make explicit. This is value that an archaeological approach may have in helping to reflect on some of our discursive practices.

3. So what? Some practical and ethical considerations

As briefly touched upon in the previous section, one common denominator of the four analytical studies is the archaeological 'insight' that our objects of knowledge are contingent and artificial constructions—products of certain discursive practices—and not the natural or necessary culmination of scientific progress. Especially the papers on SA and safety culture, the papers that rely on the most explicit Foucauldian framework, stress this point. If we are willing to buy into the archaeological arguments that situation awareness, safety culture, and resilience are (historically) contingent constructions, we implicitly acknowledge the possibility that things do not need to be, or remain, the way that they are. By recognising that some of our contemporary discursive practices are functioning counter to their explicit goals and are giving rise to some undesirable effects, it becomes an ethical choice—for us as a human factors and safety science community—to promote or denounce them. In this section, my reflections are no longer chiefly directed at a methodological or theoretical level, but now it is time to consider the implications of my archaeological 'insights' at a practical and ethical level.

In the previous section we have seen that despite human factors and safety science' commitment to a systems approach, some of their discursive practices may still be aimed at the individual as the practical target for intervention. Of course, there is nothing wrong focusing on the individual per se. It becomes problematic, both practically and ethically, when it is taken as the only approach. Explanations are always simplifications of reality—otherwise our theories and models would be as complex as the phenomena they aim to understand and would yield no explanatory or predictive value. The question is, and here it becomes an ethical issue, what knowledge or solutions do we want our human factors and safety science theories to produce? In the wake of an accident, what countermeasures do we hope to launch, what do we miss by relying on these explanations, and who wins and who loses? Are our practices serving the well-being of society or the operators that human factors claims to have in its best interest (see International Ergonomics Association, 2014)? As a human factors community, we need to ask whether the objects we propose—just as the interventions, recommendations, methods, classifications, etc.— are serving our purposes.

The finding that some of the discursive practices that human factors and safety science promote run counter to their explicit pursuit of a systems approach is not just a theoretical problem. Human factors and safety science' unknowing or implicit commitment to the individual as the major unit of analysis has a number of practical consequences. One of them, strongly exemplified by the case of Karl, may be the manner in which the causes for accidents are attributed. Whereas in the 1960's human error was said to be the main cause of accidents in about 30% of all accidents, it is now commonly accepted that 70% to 90% of accidents can be attributed to human error (Hollnagel, 2006a). Such human error explanations of accidents often see human error as largely endogenous, that is, as being caused by the characteristics inherent in humans. They see human error as a personality issue, as a motivational problem of people not trying hard enough to do their job properly, or as due to people's inherent performance limitations and cognitive shortcomings.

One further (possible) consequence of human factors and safety science' implicit commitment to self-contained individual as the major locus of intervention and explanation for accidents, might be the significant increase in the criminal prosecution of human error. Where it used to be customary in the aftermath of an accident to launch an investigation aimed solely at preventing similar accidents in the future, nowadays such investigations are almost always accompanied by judicial inquiries aimed at assessing and allocating liability (Dekker, 2009; Michaelides-Mateou & Mateou, 2010). Recent years have shown a growing concern in the safety field about a trend towards the criminalisation of professionals for not living up to their professional duties. Human factors and safety researchers and practitioners have noted these worrying trends across a wide sweep of practices, including aviation (Dekker, 2010a; Esler, 2009; Michaelides-Mateou & Mateou, 2010; Thomas, 2007), marine and shipping industries (Hed, 2005, p. 12), child protection (Alexander & Alexander, 1995), and medicine (Dekker, Sidney W. A., 2011; McDonald, 2008; Starkey & Maeda, 2009), to name just a few.

Having said all of this, the manner in which the juridical discourse takes and uses human factors objects such as SA to ask (and answer) questions about error, culpability, and criminality is an ethical problem that we as a human factors community need to take responsibility for. Instead of pointing the finger at prosecutors and juridical discourses uncritically borrowing our human factors and safety science objects and using them in 'inappropriate' manners (see chapter 2), the archaeological studies in this thesis suggest that we—as the human factors and safety science community concerned with the well being of operators in complex socio-technical systems—need to take a critical look at the discursive practices that we ourselves propose. As long as the self-contained individual remains the unit of analysis for human factors, we are asking for ethical problems in our systems; we are at not helping other discourses, such as the juridical, to move away from accounts of individuals as the fundamental elements of thought and action. We are not doing anything to point other discourses to vocabularies of complexity and systems, to (postmodern) models of causality and accountability in which systems as a whole break down, not their substituent (human) parts.

In this sense, language matters. If the human factors and safety science discourses had elected to focus on concerns with 'situational opacity', as proposed earlier in this thesis, the importance of the system—and SA's purpose of improving systems design—would have been foregrounded. The use of an object easily associated with individual mistake, lapse, and error might have been avoided. As a result, the human factors and safety science discourses themselves provide the judicial process with an epistemology and language—here, in the form of certain objects—that help it focus on the individual and her or his omissions or actions. In a way, we are thus complicit in Karl Lilgert sitting in jail today; without human factors supplying the object of SA, he might still have been in jail, but not under the current charge.

In this thesis it is not my intention to argue against the use of situation awareness, safety culture, or resilience *per se*; I am not arguing that we should drop them altogether. Rather, by providing an archaeological sketch of some prominent human factors and safety science objects, this thesis aims to show "on what type of assumptions, of familiar notions, of established, unexamined ways of thinking accepted practices are based" (Foucault, Michel 2001, p. 456). In doing so, this thesis hope to question and possibly dislocate some taken-for-granted practices that human factors and safety science promote. Invoking objects as safety culture or resilience (engineering), does not automatically provide a systems approach. As applied disciplines, we need to take the time to think about—and take responsibility for—the effects of our propositions. In line with Foucault's agenda, I hope to have argued, "together with many others, that certain phrases [such as SA, safety culture, or resilience] can no longer be spoken so lightly, certain acts no longer—or at least no longer so unhesitatingly—performed" (Foucault, 1991b, p. 83).

4. Limits of the thesis & future directions

As we have seen, the archaeological highlighting of the contingency of our knowledges provides both the need and justification for (ethical) reflection on the effect of our discursive practices. All the papers presented in this thesis conclude by expressing the need for such reflection. The last paragraph of the 'conditions of possibility for SA' paper, presented in chapter 7, for example, concludes its archaeological argument:

> By arguing for the contingent nature of our objects of knowledge, archaeology opens another agenda of inquiry. Archaeology opens an ethical discussion, asking questions about the consequences for operating a particular discursive object. … The contribution of this paper is not in making an argument about whether SA is real – or even useful – or not. It does not take a position on whether we are researching SA in an appropriate manner or not. Rather, we aim to argue that it is a historically constituted object – reinforced by science, institutions, practitioners, etc. – that emerged out of a particular logic, a certain political need, and historically available knowledges. And, as such, it is a pertinent look at the effects of our objects of knowledge. (Winsen et al., 2015, p. 65)

Compelling as such calls for (ethical) reflection about the effects of our discursive practices are, this is also where archaeology finds its limits. "Although archaeology is quite capable of describing the conceptual system underlying a practice, linguistic or not, it is not suited to describe the effects of a practice" (Gutting, 2005, p. 45). This is the largest problem that I have run into with my (purer) archaeological papers. All that an archaeological approach can convincingly do is to describe the discursive regularities that can be found in discourse, and, as such, argue for the contingent nature of our knowledges, our objects, and the discursive structures (assumptions) on which they rests. After such description, archaeology can only *call* for the need for reflection of the effects of such discursive practices (as illustrated by the quote from the conditions of possibility for SA paper, with which I started this section). It can only call for the need to look into the effects of these discursive objects. However, it does not provide the (theoretical or methodological) framework to provide such reflections itself—as such, in providing my ethical reflections on some of the practical effects of the objects studied in this thesis, the previous section might have dramatically overstepped the boundaries of an archaeological approach[24].

In other words, archaeology enables us to take apart the assumptions, knowledges, principles, and regularities that offer the foundation for any object, discourse, or discursive practice and as such provides an appropriate and convincing way to question taken-for-granted truths. It does, however, not provide a clear approach to study the

[24] However, as part of the concluding sections of a PhD thesis, I feel my (rather polemic) reflections are warranted.

effects of certain discursive constructions. The problem, as Foucault explains himself, is the "problem of causality" (Foucault, 2002b, p. xiii). He explains that in his archaeological work, "I left the problem of causes to one side, I chose instead to confine myself to [archaeological description], thinking that this would be an indispensable step if, one day, a theory of … epistemological causality was to be constructed. (Foucault, 2002b, p. xiv). This later step, towards 'a theory of epistemological causality', Foucault presents with his genealogical approach, which is concerned with the question of how our knowledges and discursive practices change.

In his genealogical work, Foucault (e.g. Foucault, 1977, 1980, 1990, 1991a) abandons the attempt to understand discourse as rule-governed systems of statements. Instead, by connecting the concepts of knowledge and power, genealogy tries to understand how different relations of power and knowledge operate to produce 'effects of truth' and 'modes of subjectivation'. Just like archaeology, genealogy is not about distinguishing true from false, nor the natural from the fabricated, but it provides a better framework for understanding how some discourses have come to be accepted as truth. In 'Discipline and Punish', for example, the focus of Foucault's genealogy is the "present scientifico-legal complex from which the power to punish derives its bases, justification, and rules" (Foucault, 1991a, p. 23). In short, with genealogy, Foucault starts to move away from the (archaeological) description of discourses 'in their own volume' and becomes interested in how power (relations) and knowledge are constitutive of one another (Foucault, 1980).

As can be observed in chapter 8—which presents both an archaeological and genealogical part to the study of safety culture as an object—the archaeological part establishes safety culture as an artificial object, which emerged under certain historically contingent conditions, and now functions with certain discursive effects. However, it is not until the genealogical part of this study, that some of the effects of this object are more explicitly discussed based on Foucault's genealogical conceptions as 'disciplinary power', 'bio-political power', and 'governmentality'. Similarly, chapter 9 shows how an implicit genealogical orientation helps to understand the effects that a 'macro' object like resilience—that claims to function at the level of organisations—may have in terms of 'subjectivation' at the micro level of the individual.

In returning to my reflections on the limitations of my study, Foucault's genealogical approach may thus be said to provide a more appropriate 'tool' for looking into the effects of our discursive practices—which constitutes a significant part of my overarching thesis question. Moreover, genealogy may also be a more appropriate tool for applied disciplines as human factors and safety science; where archaeology 'merely' provides *descriptions* of the discursive regularities that can be found in discourses and their objects, genealogy explicitly looks at their *effects*. Genealogy may thus provide more 'applied' suggestions or leverage for change, whereas archaeology stops at a call for reflection on our practices. Given my thesis' aim to bring Foucault to the disciplines of human factors and safety science, strategically, it would probably have been smarter to aim for Foucault's genealogical rather than archaeological approach—this argument can even be supported by the finding, as discussed in the introduction, that there are a few studies in the domains of human factors and safety science that draw on Foucault's genealogical approach, whereas his archaeological approach is entirely overlooked.

Does this mean, however, that Foucault's archaeological approach—which has been the heart of this thesis—to study discursive practices is useless for practically oriented disciplines as human factors and safety science? After having spent three years of my academic life with it, I am inclined to say no. In addition to my earlier argument for 'epistemological pluralism', it is important to see archaeology as just one of Foucault's 'tools' in his large emancipatory repertoire—'tool-box'. Genealogy presents another tool. However, as chapter 8 hopes to illustrate, archaeological analysis seamlessly flows into a genealogical analysis that focuses on the effects of our knowledges in terms of the power relations and subjectivities. Instead of seeing archaeology and genealogy as two distinct approaches, they can be seen as symbiotic or reciprocal. Or as Jørgensen and Ebrary (2007, p. 56) see it: "Archaeology is subsumed within genealogy; it is, in other words, subjected to genealogy's tactical purposes. Archaeology must, therefore, be discussed here since a lot of what later becomes genealogical power analysis is grounded in archaeology". Even though Foucault's genealogical approach moves away from archaeology to explore new problems—concerned with the effects of our practices—there is no rejection of the central theoretical insights of his archaeological approach. In other words, I believe it is useful to start with an archaeological understanding of the problem, before we move on to draw on genealogy to look for ways to go forward.

As such, in thinking about future ventures for study, I would like to pursue some of the archaeological 'insights' that the studies in this thesis have produced from a genealogical perspective. Where this thesis establishes that a number of objects emerge under certain conditions of possibility, an interesting next—genealogical—step could be to address the question what it means for the disciplines of human factors and safety science that their central objects of knowledge emerge under certain conditions; what are the effects of this unity in the knowledges that they promote. This question can be explored through numerous (genealogical) projects, to suggest just a few: how are alternative (human factors) knowledges or objects excluded; What power structures limit what can be legitimately said by these disciplines; What power effects and relations do these particular 'knowledges'—that focus on the self-contained individual—give rise to; What subject positions are created by these knowledges; What are the larger societal effects of these scientific knowledges, their power relations, and subjectivities. Chapter 8 might provide a good template for organising such studies, building the genealogical approach on firm archaeological grounds.

Personally, I am also keen to analyse some of the effects that our contemporary human factors and safety science objects—such as situation awareness, safety culture, or resilience—have in constituting a discourse on criminalisation. The 'negligent operator' as a particular kind of social subject, for example, was constructed, and could only make its appearance, within the moral, legal, safety and human factors discourses, practices (responses to accidents) and institutional apparatus (safety investigation boards) of the late twentieth century. As such, it would be, both theoretically and practically, interesting and (ethically) important to study how (and why) such objects are imported into the juridical discourse. Not only might such studies be able to campaign on behalf of Karl Lilgert, it might help us—as disciplines concerned with the efficiency of our socio-technological systems and the human wellbeing connected to them— ensure that human factors and safety science objects do not continue to shift the emphasis back onto individual human operators.

Another important limitation of the thesis is a direct effect of the decision to select the most cited journal articles as the 'scientific archive' for the various archaeological analyses (see chapter 5). As a result of this selection criterion, some more recent articles, concerning the objects of interest to this thesis, have been excluded, simply because these articles have not had the time to be referenced as much. As the systems approach has been gaining momentum particularly over the last decade, the various analyses presented in this thesis may have thus overlooked a part of the literature that does commit – even at an archaeological level – to a systems approach.

A recent conception of SA, 'distributed situation awareness' (DSA), for example, has been developed from a Cognitive Systems Engineering (CSE) perspective (Salmon et al., 2008; Salmon, Walker, & Stanton, 2015; Stanton et al., 2006). This perspective, rather than seeing cognition as 'in the mind', regards the environment as an active determinant of and participant in cognitive processes, as 'cognition in the wild' (see Hutchins, 1995). CSE explicitly takes the joint cognitive system (JCS) instead of individual (human) cognition as its unit of analysis, which allows for a different operationalization of SA. "Rather than trying and understanding the 'component' humans in the system by analysing their individual cognition, requiring ever more experimental complexity and effort to do so, DSA bypasses this by focussing on the interactions and 'transactions' between them" (Salmon, Walker, & Stanton, 2015, p. 180). With this focus on interaction, DSA cannot be reduced to something an individual (or even a group of individuals) possesses (a state of mind) but needs to emerge out of interactions, something a group actively constructs. Moreover, CSE not only focuses on interactions between humans, but those of the entire (joint cognitive) system—including the individual, the team, and technological artefacts: "SA is held by human and non-human agents" Stanton et al., 2006, p. 1290). In other words, "the SA of a team is distributed not only throughout the agents comprising the team, but also in the artefacts that they use in order to accomplish their goals" (Salmon et al., 2006, p. 227).

These systems conceptions of human factors objects are not entirely 'recent', as already in the seminal 1995 Human Factors special issue on situation awareness, a number of papers (to a greater or lesser extent) already argued that the value of SA lies in its ability to draw attention to the intimate relation between the situation and awareness, the world and our understanding of it (Adams et al., 1995; Flach, 1995b; Salas, 1995; Smith & Hancock, 1995). Flach, for example, argues: "This new perspective spans the human-environment system and encourages a deep appreciation for the rich interactions between human and environment and among perception, decision making, and action" (1995, p. 155). Only more recently, however, have these systems theories been gaining enough popularity (in terms of academic acceptance) to be able to compete with 'traditional' conceptions that focus on the individual. This has left these systems theories (such as CSE) largely out of the scope of this thesis' analyses. Perhaps, hopefully, if one conducts a similar analysis in a few years, the top cited journal articles will include a (much) larger number of papers that present such systems models.

I would like to end with some of my more personal reflections on the thesis as a process and final product. Obviously, some parts are better than others—particularly the extent to which the analytical papers pursue a Foucauldian, archaeological, argument varies[25]. Some of the limitations, but also the strengths, that I see in this thesis can be attributed to the realities that come with writing journal articles with others. Not only does the pursuit of journal publication (as opposed to chapters in a monograph) bring with it considerations of length, audience, and other strategic sacrifices, but also the process of writing with co-authors brings with it great joys and insights as well as the challenge to integrate various 'voices' into a single narrative. As such, I feel that my greatest accomplishment in this project is that I have managed to engage various human factors and safety science scholars to write the various archaeologically inspired papers with me. Nothing compares to the specific joy that comes with writing with other people that know what they are talking about. They pushed me to be a better writer and scholar. The total number of publications to have come out of this thesis—6 (published or submitted) papers, 1 book chapter, and then I am not counting the 3 papers that I decided to exclude from the narrative of this thesis—hopefully shows, not only my capacity to collaborate with other academics on scientific publications, but also the joy that I find in this collaborative process.

[25] I feel most insecure about chapter 9. Despite its aim for an archaeological description of the object of resilience, some parts of the paper read as a typical 'literature review' aimed at collapsing the *meaning* that the object has for various scholars. I decided to include the paper in the thesis, as, overall, I feel the argument works as an archaeological argument about the conception and effects of resilience as a object.

Reference List

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human factors, 37*(1), 85-104.

Adriaanse, L. S., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar. *The Electronic Library, 31*(6), 727-744.

Aghaei Chadegani, A., Salehi, H., Md Yunus, M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of science and scopus databases. *Asian Social Science, 9*(5), 18-26.

Alexander, E. (2013). Resilience and disaster risk reduction: An etymological journey. *Natural Hazards and Earth System Science*, *13*(11), 2707-2716.

Alexander, R., & Alexander, C. L. (1995). Criminal prosecution of child protection workers. *SOCIAL WORK, 40*(6), 809-814.

Allik, J. (2013). Factors affecting bibiometric indicators of scientific quality. *Trames Journal of the Humanities and Social Sciences, 17*(3), 199-214.

Allmark, P., Bhanbhro, S., & Chrisp, T. (2014). An argument against the focus on community resilience in public health. *BMC Public Health*, *14*, 62.

Almedom, A. M., & Glandon, D. (2007). Resilience is not the absence of PTSD any more than health is the absence of disease. *Journal of Loss and Trauma*, *12*(2), 127-143.

Almklov, P. G., Rosness, R., & Størkersen, K. (2014). When safety science meets the practitioners: Does safety science contribute to marginalization of practical knowledge? *Safety Science, 67*(0), 25-36.

Alvesson, M., & Karreman, D. (2000). Varieties of discourse: on the study of organizations through discourse analysis. *Human Relations, 5*(3), 1125-1149.

Andersen, S., & Mostue, B. A. (2012). Risk analysis and risk management approaches applied to the petroleum industry and their applicability to IO concepts. *Safety Science*, *50*(10), 2010-2019.

Antonsen, S. (2009a). Safety culture and the issue of power. *Safety Science, 47*(2), 183-191.

Antonsen, S. (2009b). Safety culture assessment – Mission impossible? On the relationship between safety and culture. *Journal of Contingencies and Crisis Management, 17*(4), 242-254.

AORN (Association of periOperative Registered Nurses). (2006). AORN Guidance Statement: Creating a Patient Safety Culture. *AORN Journal, 83*(4), 937-943.

Arribas-Ayllon, M., & Walkerdine, V. (2008). Foucauldian Discourse Analysis. In C. Willig & W. Stainton Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 91-108). Los Angeles, [Calif.]: SAGE.

Artman, H. (2000). Team situation assessment and information distribution. *Ergonomics, 43*(8), 1111-1128.

Ashby, W. R. (1956). *An introduction to cybernetics*. New York: J. Wiley.

Atak, A., & Kingma, S. (2011). Safety culture in an aircraft maintenance organisation: A view from the inside. *Safety Science, 49*, 268-278.

Australian Transportation Safety Bureau. (2007). *CFIT: Australia in context 1996 to 2005*. Canberra, ACT: ATSB.

Aven, T. (2014). What is safety science? *Safety Science, 67*(0), 15-20.

Aviation Safety Council. (2002). *Crashed on a Partially Closed Runway during Takeoff, Singapore Airlines Flight 006, Boeing 747-400, 9V-SPK, CKS Airport, Taoyuan, Taiwan, October 31, 2000*. (ASC-AAR-02-04-001). Taipei, Taiwan: ACS.

Azadeh, A., Salehi, V., Arvan, M., & Dolatkhah, M. (2014). Assessment of resilience engineering factors in high-risk environments by fuzzy cognitive maps: A petrochemical plant. *Safety Science*, *68*, 99-107.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Ball, S. J. (1995). Intellectuals or Technicians? The Urgent Role of Theory in Educational Studies. *British Journal of Educational Studies*(3), 255.

Bakx, G. C., & Nyce, J. M. (2012). Is redundancy enough?: A preliminary study of apache crew behaviour. *Theoretical Issues in Ergonomics Science*, 1-15.

Barker, J. R., & Cheney, G. (1994). The Concept and the Practices of Discipline in Contemporary Organizational Life. *COMMUNICATION MONOGRAPHS, 61*(1), 19-43.

Barratt, E. (2008). The later Foucault in organization and management studies. *Human Relations 61*, 515–537.

Batteau, A. W. (2001). The anthropology of aviation and flight safety. *Human organization, 60*(3), 201-211.

Batteau, A. W. (2002). Anthropological approaches to culture, aviation, and flight safety. *Human Factors and Aerospace Safety, 2*(2), 147-171.

Beach, C., Croskerry, P., & Shapiro, M. (2003). Profiles in patient safety: Emergency care transitions. *Academic Emergency Medicine, 10*(4), 364-367.

Bedny, G., & Meister, D. (1999). Theory of Activity and Situation Awareness. *International Journal of Cognitive Ergonomics, 3*(1), 63-72.

Benn, J., Healey, A. N., & Hollnagel, E. (2008). Improving performance reliability in surgical systems. *Cognition, Technology & Work*, *10*(4), 323-333.

Bergström, O., Hasselbladh, H., Kärreman, D., & La Porte, T. R. (2009). Organizing disciplinary power in a knowledge organization. *Scandinavian Journal of Management, 25*(2), 178-190.

Bergström, O., & Knights, D. (2006). Organizational discourse and subjectivity: Subjectification during processes of recruitment. *Human Relations, 59*(3), 351-377.

Billings, C. E. (1995). *Situation Awareness Measurement and Analysis: A Commentary.* Paper presented at the International Conference on Experimental Analysis and Measurement of Situation Awareness.

Blandford, A., & Wong, B. L. W. (2004). Situation awareness in emergency medical dispatch. *International Journal of Human Computer Studies, 61*(4), 421-452.

Boin, A., & Schulman, P. (2008). Assessing NASA's Safety Culture: The limits and possibilities of high-reliability theory. *Public Administration Review, 68*(6), 1050-1062.

Bradbury-Jones, C., Irvine, F., & Sambrook, S. (2007). Unity and detachment: a discourse analysis of doctoral supervision. *International journal of qualitative methods, 6*(4), 81-96.

Bradley, C. A., Rolka, H., Walker, D., & Loonsk, J. (2005). BioSense: implementation of a National Early Event Detection and Situational Awareness System. *MMWR. Morbidity and mortality weekly report, 54 Suppl*, 11-19.

Breakey, H. (2009). The Epistemic and Informational Requirements of Utilitarianism. *Utilitas, 21*(1), 72-99.

Bronstein, A. C., Spyker, D. A., Cantilena Jr, L. R., Green, J. L., Rumack, B. H., & Giffin, S. L. (2009). 2008 Annual report of the American association of poison control centers' national poison data system (NPDS): 26th annual report AAPCC 2009 annual report of the NPDS. *Clinical Toxicology, 47*(10), 911-1084.

Bronstein, A. C., Spyker, D. A., Cantilena Jr, L. R., Rumack, B. H., & Dart, R. C. (2012). 2011 Annual report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 29th Annual Report. *Clinical toxicology (Philadelphia, Pa.), 50*(10), 911-1164.

Brooker, P. (2010). SESAR safety decision-making: Lessons from environmental, nuclear and defense modeling. *Safety Science*, *48*(7), 831-844.

Brown, A. D. (2004). Authoritative Sensemaking in a Public Inquiry Report. *Organization Studies, 25*(1), 95-112.

Bruxelles, S. (2010). Coroner criticises US as he gives 'friendly fire' inquest verdict. *The Times*. Retrieved from http://www.thetimes.co.uk/tto/news/world/asia/afghanistan/article249300.ece

Bruyelle, J. -L., O'Neill, C., El-Koursi, E. -M., Hamelin, F., Sartori, N., & Khoudour, L. (2014). Improving the resilience of metro vehicle and passengers for an effective emergency response to terrorist attacks. *Safety Science*, *62*, 37-45.

Burke, J. L., Murphy, R. R., Coover, M. D., & Riddle, D. L. (2004). Moonlight in Miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. *Human-Computer Interaction, 19*(1-2), 85-116.

CAIB. (2003). *Final Report on Columbia Space Shuttle Accident*. Washington, DC: Government Printing Office.

Campbell, E. M., Guappone, K. P., Sittig, D. F., Dykstra, R. H., & Ash, J. S. (2009). Computerized provider order entry adoption: Implications for clinical workflow. *Journal of General Internal Medicine, 24*(1), 21-26.

Capra, F. (1982). *The turning point: science, society, and the rising culture*. New York: Simon and Schuster.

Capra, F. (1996). *The web of life: a new synthesis of mind and matter*. London: HarperCollins.

Carmeli, A., Friedman, Y., & Tishler, A. (2013). Cultivating a resilient top management team: The importance of relational connections and strategic decision comprehensiveness. *Safety Science*, *51*(1), 148-159.

de Carvalho, P. V. R. (2011). The use of functional resonance analysis method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. *Reliability Engineering & System Safety*, *96*(11), 1482-1498.

Catchpole, K., Mishra, A., Handa, A., & McCulloch, P. (2008). Teamwork and error in the operating room: Analysis of skills and roles. *Annals of Surgery, 247*(4), 699-706.

CENIPA. (2008). *Aeronautical Accident, PR–GTD and N600XL, B-737 8EH and EMB-135 BJ LEGACY, 29 September 2006*. (A-00X/CENIPA/2008). Brasilia: CENIPA.

Chan, A. (2000). Redirecting Critique in Postmodern Organization Studies: The Perspective of Foucault. *Organization Studies, 21*(6), 1059-1075.

Chapanis, A. (1999). *The Chapanis chronicles : 50 years of human factors research, education and design*. Santa Barbara, Calif.: Aegean.

Chen, J. Y. C., Haas, E., & Barnes, M. J. (2007). Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 37*(6), 1231-1245.

Chikudate, N. (2009). If human errors are assumed as crimes in a safety culture: A life world analysis of a rail crash. *Human Relations, 62*(9), 1267-1287.

Choi, H. L., Brunet, L., & How, J. P. (2009). Consensus-based decentralized auctions for robust task allocation. *IEEE Transactions on Robotics, 25*(4), 912-926.

Choudhry, R.M., Fang, D., & Mohamed, S. (2007). The nature of safety culture: a survey of the state-of-the-art. *Safety Science, 45*, 993-1012.

Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems.* London: Routledge.

Clarke, S. (2000). Safety culture: under-specified and overrated? *International Journal of Management Reviews, 2*(1), 65-90.

Clegg, S. R. (1989). Radical Revisions: Power, Discipline and Organizations. *Organization Studies, 10*(1), 97-115.

Cohen, A. (1977). Factors in successful occupational safety programs. *Journal of Safety Research, 9*, 4, 168-178.

Cohen, A., Smith, M, & Cohen, H.H. (1975). Safety Program Practices in High vs. Low Accident rate Companies – an Interim Report (US Department of Health, Education and Welfare Publication 75-185). Cincinnati, OH: National Institute for Occupational Health and Safety.

Colla, J. B., Bracken, A. C., Kinney, L. M., & Weeks, W. B. (2005). Measuring patient safety climate: A review of surveys. *Quality & Safety in Health Care, 14*(5), 364-366.

Collins, J. A., & Fauser, B. C. J. M. (2004). Balancing the strengths of systematic and narrative reviews. *Human reproduction update, 11*(2), 103-104.

Collins, R. T., Lipton, A. J., Fujiyoshi, H., & Kanade, T. (2001). Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE, 89*(10), 1456-1477.

Collinson, D.L. (1998). "Shifting lives": work-home pressure in North Sea oil industry. *The Canadian Review of Sociology and Anthropology, 35*, 21-44.

Collinson, D.L. (1999). "Surviving the Rigs": Safety and Surveillance on North Sea Oil Installations. *Organisation Studies, 20*, 579-600.

Collis, L., Schmid, F., & Tobias, A. (2013). Managing incidents in a complex system: A railway case study. *Cognition, Technology & Work.*

Cook, R. I., & Nemeth, C. P. (2010). Those found responsible have been sacked": some observations on the usefulness of error. *Cognition, Technology & Work, 12*(2), 87-93.

Cook, R. I., & Woods, D. D. (1996). Adapting to new technology in the operating room. *Human factors, 38*(4), 593-613.

Cooper, M.D. (2000). Towards a model of safety culture. *Safety Science, 36*, 111-136.

Cooper, M.D., & Phillips, R.A. (1995). Killing two birds with one stone: achieving quality via total quality via total safety management. *Leadership and Organizational Development Journal, 16*, 3-9.

Corcoran, M. (2000). What is negligence? *BJU international, 86*(3), 280-285.

Corlett, E. N. (2000). Ergonomics and ethics in a changing society. *Applied ergonomics, 31*(6), 679-683.

Cornelissen, M., Salmon, P. M., Jenkins, D. P., & Lenné, M. G. (2013). A structured approach to the strategies analysis phase of cognitive work analysis. *Theoretical Issues in Ergonomics Science*, *14*(6), 546-564.

Costella, M. F., Saurin, T. A., & de Macedo Guimarães, L. B. (2009). A method for assessing health and safety management systems from the resilience engineering perspective. *Safety Science*, *47*(8), 1056-1067.

Coury, B. G., Ellingstad, V. S., & Kolly, J. M. (2010). Transportation Accident Investigation: The Development of Human Factors Research and Practice. *Reviews of Human Factors and Ergonomics, 6*(1), 1-33.

Cox, S.J., & Flin, R. (1998). Safety culture: philosopher's stone or a man of straw? *Work and Stress, 12*(3), 189-201.

Cummings, M. L. (2004). The need for command and control instant message adaptive interfaces: Lessons learned from Tactical Tomahawk human-in-the-loop simulations. *Cyberpsychology and Behavior, 7*(6), 653-661.

de Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics, 73*(1), 53-78.

Deal, T.E., & Kennedy, A.A. (1982). *Corporate Cultures: The rites and rituals of corporate life*. Reading, MA: Addison-Wesley.

Dekker, S., Nyce, J. M., Winsen, R. v., & Henriqson, E. (2010). Epistemological Self-Confidence in Human Factors Research.

Dekker, S. W. A. (2001). The re-invention of human error. *Human Factors and Aerospace Safety, 1*(3), 247-265.

Dekker, S. W. A. (2002a). *The field guide to human error investigations*. Aldershot, England; Burlington, VT: Ashgate.

Dekker, S. W. A. (2002b). Reconstructing human contributions to accidents: the new view on error and performance. *Journal of Safety Research, 33*(3), 371-385.

Dekker, S. W. A. (2004). Why we need new accident models. *Human Factors and Aerospace Safety*, 1-18.

Dekker, S. W. A. (2005). *Ten questions about human error*. Mahwah, N.J.: Lawrence Erlbaum associates.

Dekker, S. W. A. (2006). *The field guide to understanding human error*. Aldershot, England; Burlington, VT: Ashgate.

Dekker, S. W. A. (2009). Prosecuting professional mistake: Secondary victimization and a research agenda for criminology. *International Journal of Criminal Justice Sciences, 4*(1), 60.

Dekker, S. W. A. (2010a). Pilots, Controllers and Mechanics on Trial: Cases, Concerns and Countermeasures. *International Journal of Applied Aviation Studies, 10*(Journal Article), 31-49.

Dekker, S. W. A. (2010b). We have Newton on a retainer: reductionism when we need systems thinking. *Joint Commission journal on quality and patient safety / Joint Commission Resources, 36*(4), 147.

Dekker, S. W. A. (2011). The criminalization of human error in aviation and healthcare: A review. *Safety Science, 49*(2), 121-127.

Dekker, S. W. A. (2011). *Drift into failure: from hunting broken components to understanding complex systems*. Farnham, England; Burlington, VT: Ashgate.

Dekker, S. W. A. (2012a). On the epistemology and ethics of communicating a Cartesian consciousness. *Safety Science, 56*, 96-99.

Dekker, S.W.A. (2012b). *Just Culture: Balancing Safety and Accountability* (2nd Ed.). Aldershot, UK: Ashgate.

Dekker, S. W. A. (2014). The psychology of accident investigation: epistemological, preventive, moral and existential meaning-making. *Theoretical Issues in Ergonomics Science Theoretical Issues in Ergonomics Science*(1), 1-12.

Dekker, S. W. A., Bergström, J., Amer-Wåhlin, I., & Cilliers, P. (2013). Complicated, complex, and compliant: Best practice in obstetrics. *Cognition, Technology & Work*, *15*(2).

Dekker, S. W. A., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work, 6*(2), 79-86.

Dekker, S. W. A., Hummerdal, D. H., & Smith, K. (2010). Situation awareness: some remaining questions. *Theoretical Issues in Ergonomics Science, 11*, 131-135.

Dekker, S. W. A., & Lutzhoft, M. (2004). Correspondence, Cognition and Sensemaking: A Radical Empiricist View of Situation Awareness. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: theory and application* (pp. 22-41). Burlington, VT: Ashgate.

Dekker, S., & Pruchnicki, S. (2013). Drifting into failure: Theorising the dynamics of disaster incubation. *Theoretical Issues in Ergonomics Science*, 1-11.

Deninson, D.R. (1996). What is the difference between organization culture and organization climate? A native's point of view on a decade of paradigms war. *Academy of Management Review, 21*, 619-654.

Derrida, J. (1978). *Writing and difference*. London: Routledge and Kegan Paul.

Derrida, J. (1998). *Of Grammatology* (G. C. Spivak, Trans.). Baltimore: The John Hopkins University Press.

Dörner, D. (1996). *The logic of failure: recognizing and avoiding error in complex situations*. Cambridge, Mass.: Perseus Books.

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and Cell Phone Conversations in Simulated Driving. *Journal of Experimental Psychology: Applied, 14*(4), 392-400.

Dreyfus, H. L., & Rabinow, P. (1982). *Michel Foucault: beyond structuralism and hermeneutics*. New York: Harvester.

Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W. S., . . . van der Doelen, B. (2012). A strategy for human factors/ergonomics: developing the discipline and profession. *Ergonomics, 55*(4), 377-395.

Durieux, V., & Gevenois, P. A. (2010). Bibliometric indicators: quality measurements of scientific publication. *Radiology, 255*(2), 342-351.

Durso, F. T., & Gronlund, S. D. (1999). Situation Awareness. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindays & M. T. H. Chi (Eds.), *Handbook of applied cognition* (pp. 283-314). Chichester: Wiley.

Edwards, J.R.D., & Jabs, L.B. (2009). When safety culture backfires: unintended consequences of half-shared governance in a high tech workplace. *Safety Science, 46*, 707-723.

Ek, A., et al. (2007). Safety Culture in Swedish air traffic control. *Safety Science, 45*, 791-811.

Elsevier. (2014). Scopus, facts and figures. from http://www.elsevier.com/online-tools/scopus

Endsley. (2014). SA: Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making*.

Endsley, M. R. (1988). *Situation Awareness Global Assessment Technique (SAGAT)*. Paper presented at the National Aerospace and Electronics Conference, New York.

Endsley, M. R. (1995a). Measurement of Situation Awareness in Dynamic Systems. *Human factors, 37*(1), 65-84.

Endsley, M. R. (1995b). Toward a Theory of Situation Awareness in Dynamic Systems. *Human factors, 37*(1), 32-64.

Endsley, M. R. (2000). Theoretical Underpinnings of Situation Awareness: A Critical Review. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness: Analysis and Measurement* (pp. 3 - 32). London: L. Erlbaum Associates.

Endsley, M. R., Bolte, B., & Jones, D. G. (2003). *Designing for situation awareness: an approach to user-centered design*. London; New York: Taylor & Francis.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462-492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors, 37*(2), 381-394.

Esbester, M. (2008). Organizing work: company magazines and the discipline of safety. *Management & organizational history, 3*(3/4), 217-237.

Esler, D. (2009). Flight risk: the threat of criminalization. *Business & Commercial Aviation, 104*(3), 28-38.

Eva, K. W., & Regehr, G. (2007). Knowing when to look it up: A new conception of self-assessment ability. *Academic Medicine, 82*(10 SUPPL.), S81-S84.

Evans, B., & Reid, J. (2013). Dangerously exposed: The life and death of the resilient subject. *Resilience*, *1*(2), 83-98.

Ewald, F. (2002). The Return of Descartes's Malicious Demon: An Outline of a Philosophy of Precaution. In T. Baker & J. Simon (Eds.), *Embracing risk : the* *changing culture of insurance and responsibility* (pp. 273-301). Chicago: University of Chicago Press.

Feldman, R.M., & Feldman, S.P. (2006). What links the chain: an essay on organizational remembering as practice. *Organization, 13*(6), 861-887.

Fennell, D. (1988). Investigation into the King's Cross Underground Fire. London: HMSO.

Fermández-Muñiz, B., Montes-Péon, J., & Vázquez-Ordás, C.J. (2007). Safety culture: Analysis of the casual relationships between its key dimensions. *Journal of Safety Research, 38*, 627-641.

Fitts, P. M. (1947). *Psychological research on equipment design*. Washington: U.S. Govt. Print. Off.

Fitts, P. M., & Jones, R. E. (1947). Analysis of factors contributing to 460 'pilot error' experiences in operating aircraft controls. Wright-Patterson Air Force Base, Dayton, Ohio: Aero Medical Laboratory, Air Material Command.

Flach, F. F. (1988). *Resilience: Discovering a new strength at times of stress.* Fawcett Columbine New York.

Flach, J. M. (1995a). *Maintaining Situation Awareness when Stalking Cognition in the Wild.* Paper presented at the International Conference on Experimental Analysis and Measurement of Situation Awareness.

Flach, J. M. (1995b). Situation Awareness: Proceed with Caution. *Human factors, 37*(1), 149-157.

Flach, J. M. (1996). Situation Awareness: In Search of Meaning. *Gateway, 6*(6), 1-4.

Fletcher, G., Flin, R., McGeorge, M., Glavin, R., Maran, N., & Patey, R. (2003). Anaesthetists' non-technical skills (ANTS): Evaluation of a behavioural marker system. *British Journal of Anaesthesia, 90*(5), 580-588.

Flin, R., et al. (2000). Measuring the safety climate: identifying the common features. *Safety Science, 34*, 177-192.

Flin, R., Yule, S., Paterson-Brown, S., Maran, N., Rowley, D., & Youngson, G. (2007). Teaching surgeons about non-technical skills. *Surgeon, 5*(2), 86-89.

Foucault, M. (1977). Nietzsche, Genealogy, History. In M. Foucault (Ed.), *Language, counter-memory, practice: selected essays and interviews* (pp. 139-164). Ithaca, N.Y: Cornell University Press.

Foucault, M. (1980). *Power/knowledge: selected interviews and other writings, 1972-1977* (C. Gordon, Trans.). New York: Pantheon Books.

Foucault, M. (1981). The Order of the Discourse. In R. Young (Ed.), *Untying the text: a post-structuralist reader* (pp. 51-78). Boston: Routledge & Kegan Paul.

Foucault, M. (1990). *The history of sexuality. Volume 1, An introduction*. New York: Vintage.

Foucault, M. (1991a). *Discipline and punish: the birth of the prison*. Ringwood, Vic: Penguin Books.

Foucault, M. (1991b). Questions of method. In G. Burchell, C. Gordon & P. H. Miller (Eds.), *The Foucault effect: studies in governmentality* (pp. 73 - 86). Chicago: University of Chicago Press.

Foucault, M. (2001a). Essential works of Michel Foucault, Volume III: Power In J. D. Faubion (Ed.), *The essential works of Michel Foucault, Volume III, Power* (pp. 239-297). New York; New York: New Press.

Foucault, M. (2001b). *Madness and Civilization: A History of Insanity in the Age of Reason*. London: Routledge.

Foucault, M. (2001). So is it important to think? In J. D. Faubion (Ed.), *Essential works of Michel Foucault, Volume III: Power* (pp. 454-458). New York: New Press.

Foucault, M. (2002a). *The Archeology of Knowledge*. London: Routledge.

Foucault, M. (2002b). *The order of things: an archaeology of the human sciences*. London: Routledge.

Foucault, M. (2004). *Security, Territory, Population: Lectures at the Collège de France 1977-1978*. New York: Picador.

Foucault, M. (2008). *The Birth of Biopolitics: Lectures at the Collège de France 1978-1979*. Basingstoke: Palgrave Macmillan.

Foucault Studies. (2015). Submissions: Author guidelines. Retrieved from http://ej.lib.cbs.dk/index.php/foucault-studies/about/submissions - authorGuidelines

Fraizer, C.B., et al. (2013). A hierarchical factor analysis of a safety culture survey. *Safety Science, 45*, 15-28.

France, D. J., Levin, S., Hemphill, R., Chen, K., Rickard, D., Makowski, R., . . . Aronsky, D. (2005). Emergency physicians' behaviors and workload in the presence of an electronic whiteboard. *International Journal of Medical Informatics, 74*(10), 827-837.

Francis, R., & Bekera, B. (2014). A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety, 121*, 90-103.

Furniss, D., Back, J., Blandford, A., Hildebrandt, M., & Broberg, H. (2011). A resilience markers framework for small teams. *Reliability Engineering & System Safety*, *96*(1), 2-10.

Gaba, D. M., Howard, S. K., & Small, S. D. (1995). Situation Awareness in Anesthesiology. *Human factors, 37*(1), 20-31.

GAIN (Global Aviation Information Network). (2002). *Operator's Flight Safety Handbook*. Available at: http://flightsafety.org/files/OFSH_english.pdf

Garrity, Z. (2010). Discourse Analysis, Foucault and Social Work Research: Identifying Some Methodological Complexities. *Journal of Social Work, 10*(2), 193-210.

Gephart, R.P. (2004). Normal risk: Technology, sense making, and environmental disasters. *Organization & Environment, 17*(1), 20-26.

Gephart, R.P., Van Maanen, J., & Oberlechner, T. (2009). Organisations and Risks in Late Modernity. *Organisation Studies, 30*, 141-155.

Gergen, K. J. (1999). *An invitation to social construction*. London: Sage.

Gherardi, S., & Nicolini, D. (2000). To transfer is to transform: The circulation of safety knowledge. *Organization, 7*(2), 329-48.

Gherardi, S., Nicolini, D., & Odella, F. (1998). What do you mean by safety? Conflicting perspectives on accident causation and safety management in a construction firm. *Journal of Contingencies and Crisis Management, 6*, 202-213.

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher, 5*(10), 3-8.

Glendon, A.I., & Stanton, N.A. (2000). Perspectives in safety culture. *Safety science, 34*(1), 193-213.

Golden-Biddle, K., & Locke, K. (1993). Appealing Work: An Investigation of How Ethnographic Texts Convince. *Organization Science, 4*(4), 595-616.

Gomes, J. O., Borges, M. R., Huber, G. J., & Carvalho, P. V. (2014). Analysis of the resilience of team performance during a nuclear emergency response exercise. *Applied Ergonomics*, *45*(3), 780-8.

Gomes, J. O., Woods, D. D., Carvalho, P. V., Huber, G. J., & Borges, M. R. (2009). Resilience and brittleness in the offshore helicopter transportation system: The identification of constraints and sacrifice decisions in pilots' work. *Reliability Engineering & System Safety*, *94*(2), 311-319.

Gonaçalves Filho, A.P., Andrade, J.C.S., & Marinho, M.M.O. (2010). A safety culture maturity model for petrochemical companies in Brazil*. Safety Science, 48*, 615-624.

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49*(12-13), 1312-1325.

Gorman, P., Ash, J., Lavelle, M., Lyman, J., Delcambre, L., Maier, D., . . . Bowers, S. (2000). Bundles in the wild: Managing information to solve problems and maintain situation awareness. *Library Trends, 49*(2), 266-289.

Graham, L. J. (2005). *Discourse analysis and the critical use of Foucault*. Paper presented at the Australian Association for Research in Education, Sydney.

Graham, L. J. (2011). The Product of Text and 'Other' Statements: Discourse analysis and the critical use of Foucault. *Educational Philosophy and Theory, 43*(6), 663-674.

Gray, G.C. (2009). The responsibilization strategy of health and safety: Neo-liberalism and the reconfiguration of individual responsibility for risk. *British Journal of Criminology, 49*(3), 326-342.

Green, J. (2003). The ultimate challenge for risk technologies: Controlling the accidental. In J. Summerton & B. Berner (Eds.), *Constructing risk and safety in technological practice*. London: Routledge.

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology, 5*, 347-358.

Griffith University. (2014). Thesis as a series of published and unpublished papers. Retrieved January 24, 2015, from http://www.griffith.edu.au/higher-degrees-research/current-research-students/thesis/preparation/thesis-as-a-series-of-published-unpublished-papers

Grote, G., Weichbrodt, J. C., Günter, H., Zala-Mezö, E., & Künzle, B. (2009). Coordination in high-risk organizations: The need for flexible routines. *Cognition, Technology & Work*, *11*(1), 17-27.

Guarnieri, M. (1992). Landmarks in the History of Safety. *Journal of Safety Research, 23*(3), 151-158.

Guerlain, S., Adams, R. B., Turrentine, F. B., Shin, T., Guo, H., Collins, S. R., & Calland, J. F. (2005). Assessing team performance in the operating room: Development and use of a "black-box" recorder and other tools for the intraoperative environment. *Journal of the American College of Surgeons, 200*(1), 29-37.

Gugerty, L. J. (1997). Situation Awareness during Driving: Explicit and Implicit Knowledge in Dynamic Spatial Memory. *Journal of Experimental Psychology: Applied, 3*(1), 42-66.

Guldenmund, F. W. (2000). The nature of safety culture: a review of theory and research. *Safety Science, 34*(1), 215–257.

Gurses, A. P., & Xiao, Y. (2006). A Systematic Review of the Literature on Multidisciplinary Rounds to Design Information Technology. *Journal of the American Medical Informatics Association, 13*(3), 267-276.

Gutting, G. (1989). *Michel Foucault's archaeology of scientific reason*. Cambridge [England]; New York: Cambridge University Press.

Gutting, G. (2005). *Foucault: a very short introduction*. New York: Oxford University Press.

Gutwin, C., & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work: CSCW: An International Journal, 11*(3-4), 411-446.

Haavik, T. K. (2014). On the ontology of safety. *Safety science., 67*, 37-43.

Hale, A. R. (2004). Editorial: culture's confusions. *Safety Science, 34*, 1-14.

Hale, A. R. (2014). Foundations of safety science: A postscript. *Safety Science, 67*(0), 64-69.

Hale, A. R., & Heijer, T. (2006). Defining resilience. In E. Hollnagel, D. D. Woods, & N. Leveson (Eds.), *Resilience engineering: Concepts and precepts* (pp. 35-40). Aldershot: Ashgate Publishing Limited.

Hale, A. R., & Hovden, J. (1998). Management and Culture: the third age of safety. A review of approaches to organisational aspects of safety, health and environment. In. A.M. Feyer, A. Williamson (Eds.), *Occupational injury: Risk prevention and intervention*. London, UK: Taylor & Francis.

Hall, S. (2001). Foucault: Power, Knowledge and Discourse. In M. Wetherell, S. Yates & S. Taylor (Eds.), *Discourse theory and practice: a reader* (pp. 72-81). Thousand Oaks, [Calif.]: SAGE.

Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., . . . Tian, Y. L. (2005). Smart video surveillance. *IEEE Signal Processing Magazine, 22*(2), 38-51.

Hancock, P. A., & Drury, C. G. (2011). Does human factors/ergonomics contribute to the quality of life? *Theoretical Issues in Ergonomics Science, 12*(5), 416-426.

Haukelid, K. (2008). Theories of (safety) culture revisited: An anthropological approach. *Safety Science, 46*, 413-426.

Hauser, C. H., Bakken, D. E., & Bose, A. (2005). A failure to communicate. *IEEE Power and Energy Magazine, 3*(2), 47-55.

Healy, S. (2003). Epistemological pluralism and the 'politics of choice'. *Futures, 35*(7), 689-701.

Hed, B. (2005). Criminalisation of Seafarers - will this contribute to improving the safety and environment at sea, or could it in fact be counterproductive? *The Swedish Club Letter, 1,* 12-15.

Heinrich, H. W. (1941). *Industrial accident prevention; a scientific approach*. New York and London: McGraw-Hill Book Company, Inc.

Heinrich, H. W., Petersen, D. P. E., & Roos, D. B. A. (1980). *Industrial accident prevention: a safety management approach*. New York: McGraw-Hill.

Helmreich, R.L. (2001). On error management: lessons from aviation. *British Medical Journal, 320*, 781-785.

Helmreich, R. L., & Davies, J. M. (1996). Human factors in the operating room: Interpersonal determinants of safety, efficiency and morale. *Bailliere's Clinical Anaesthesiology, 10*(2), 277-295.

Helton, W. S., & Kemp, S. (2011). What basicapplied issue? *Theoretical Issues in Ergonomics Science, 12*(5), 397-407.

Henriques, J., Hollway, W., Urwinn, C., Venn, C., & Walkerdine, V. (2002). Selections from Changing the Subject: Psychology, Social Regulation and Subjectivity. *Feminism & Psychology, 12*(4), 427-431.

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research, 37*(4), 401-411.

Hoffman, R. R., Marx, M., Amin, R., & McDermott, P. L. (2010). Measurement for evaluating the learnability and resilience of methods of cognitive work. *Theoretical Issues in Ergonomics Science*, *11*(6), 561-575.

Hogg, D. N., Folleso, K., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics, 38*(11), 2394-2413.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 1-23.

Hollnagel, E. (2004). *Barriers and accident prevention*. Aldershot, Hampshire, England; Burlington, VT: Ashgate.

Hollnagel, E. (2005). Human Reliability Assessment in Context. *Nuclear Engineering and Technology, 37*, 159-166.

Hollnagel, E. (2006a). Accident Analysis and "Human Error". In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 1889-1892). Boca Raton, FL: Taylor & Francis.

Hollnagel, E. (2006b). Resilience - the challenge of the unstable. In E. Hollnagel, D. Woods, & N. Leveson (Eds.), *Resilience engineering, concepts and precepts* (pp. 9-17). Aldershot: Ashgate Publishing Company.

Hollnagel, E. (2008). Prologue: The scope of resilience engineering. In E. Hollnagel, J. Pariès, D. D. Woods, & J. Wrethall (Eds.), *Resilience engineering in practice - A guidebook* (pp. xxix-xxxix). Farnham: Ashgate Publishing Limited.

Hollnagel, E. (2009). *The ETTO Principle-- efficiency-thoroughness trade-off : why things that go right sometimes go wrong*. Farnham, England: Ashgate.

Hollnagel, E. (2014a). Human factors/ergonomics as a systems discipline? ''The human use of human beings'' revisited. *Applied ergonomics, 45*(1), 40-44.

Hollnagel, E. (2014b). *Safety-I and Safety-II: The Past and Future of Safety Management*: Ashgate Publishing Ltd.

Hollnagel, E. (2014c). Is safety a subject for science? *Safety Science, 67*(0), 21-24.

Hollnagel, E., & Amalberti, R. (2001). The emperor's new clothes: Or whatever happened to 'human error'? *Paper presented at the 4th international workshop on human error, safety and systems development*, Linköping, Sweden.

Hollnagel, E., & Fujita, Y. (2013). The Fukushima Disaster – Systemic Failures as the lack of Resilience. *Nuclear Engineering and Technology*, *45*(1), 13-20.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: foundations of cognitive systems engineering*. Boca Raton, FL: Taylor & Francis.

Hollnagel, E., Woods, D., & Leveson, N. (2006a). *Resilience engineering, concepts and precepts.* Aldershot: Ashgate Publishing Company.

Hollnagel, E., Woods, D. D., & Leveson, N. (2006b). Prologue: Resilience engineering concepts. In E. Hollnagel, D. D. Woods, & N. Leveson (Eds.), *Resilience engineering: Concepts and precepts* (pp. 1-6). Aldershot: Ashgate Publishing.

Holmes, D., Murray, S. J., Perron, A., & McCabe, J. (2008). Nursing Best Practice Guidelines: reflecting on the obscene rise of the void. *JONM Journal of Nursing Management, 16*(4), 394-403.

Holmes, D., Murray, S. J., Perron, A. l., & Rail, G. v. (2006). Deconstructing the evidence-based discourse in health sciences. *International Journal of Evidence-Based Healthcare International Journal of Evidence-Based Healthcare, 4*(3), 180-186.

Hook, D. (2001). Discourse, Knowledge, Materiality, History: Foucault and Discourse Analysis. *Theory & Psychology, 11*(4), 521-547.

Hopkins, A. (2006). Studying organisational cultures and their effects on safety. *Safety Science, 44*, 875-889.

Hopkins, A. (2013). Issues in safety science. *Safety Science.*

Hornborg, A. (2013). Revelations of resilience: From the ideological disarmament of disaster to the revolutionary implications of (p)anarchy. *Resilience*, *1*(2), 116-129.

Horswill, M. S., & McKenna, F. P. (2004). Driver's Hazard Perception Ability: Situation Awareness. In S. Banbury & S. b. Tremblay (Eds.), *A cognitive approach to situation awareness: theory and application* (pp. 155-175). Burlington, VT: Ashgate.

Howarth, D. (2000). *Discourse*. Philadelphia: Open University Press.

Hsieh, M. A., Cowley, A., Keller, J. E., Chaimowicz, L., Grocholsky, B., Kumar, V., . . . MacKenzie, D. C. (2007). Adaptive teams of autonomous aerial and ground Robots for situational awareness. *Journal of Field Robotics, 24*(11-12), 991-1014.

Hsieh, M. A., Cowley, A., Kumar, V., & Taylor, C. J. (2008). Maintaining network connectivity and performance in robot teams. *Journal of Field Robotics, 25*(1-2), 111-131.

Hudson, P. (2007). Implementing safety culture in a major multi-national. *Safety Science, 45*(6), 697-722.

Hugh, T. B., & Dekker, S. W. A. (2009). Hindsight bias and outcome bias in the social construction of medical negligence: a review. *Journal of Law and Medicine, 16*(5), 846-857.

Humes, W., & Bryce, T. (2003). Post-structuralism and policy research in education. *JOURNAL OF EDUCATION POLICY, 18*(2), 175-187.

Hummerdal, D. H., Wilhemsson, A., & Dekker, S. W. A. (2013). Learning from failure. In J. D. Lee & A. Kirlik (Eds.), *The Oxford handbook of cognitive engineering* (pp. 404-412). New York, NY: Oxford University Press.

IAEA (International Atomic Energy Agency). (1998). *Developing safety culture in nuclear activities: practical suggestions to assist progress*. (Safety Report Series No 11). Vienna, Austria: IAEA.

IAEA (International Atomic Energy Agency) (2002a). *Safety culture in nuclear installations: guidance for use in the enhancement of safety culture* (IAEA-TECDOC-1329). Vienna, Austria: IAEA.

IAEA (International Atomic Energy Agency). (2002b). *Self-assessment of safety culture in nuclear installations: highlights and good practices* (IAEA-TECDOC-1321). Vienna, Austria: IAEA.

ICAO (International Civil Aviation Organisation). (2009). *Safety Management Manual (SMM)* (2nd ed.). Montreal, Canada: ICAO.

INSAG (International Nuclear Safety Advisory Group). (1986). *Summary Report on the Post-Accident Review Meeting on the Chernobyl Accident*. (Safety Series No. 75-INSAG-1). International Atomic Energy Agency, Vienna.

INSAG (International Nuclear Safety Advisory Group). (1988). *Basic Safety Principles for Nuclear Power Plants* (Safety Series No. 75-INSAG-3). International Atomic Energy Agency, Vienna.

INSAG (International Nuclear Safety Advisory Group). (1991). *Safety Culture* (Safety Series No. 75-INSAG-4). International Atomic Energy Agency, Vienna.

International Civil Aviation Organization, Flight Safety Foundation, & Federal Aviation Administration. (1995). *Controlled flight into terrain: education and training aid*. Alexandria, VA: Flight Safety Foundation.

International Ergonomics Association. (2014). Definition and Domains of ergonomics. Retrieved April 19, 2014, 2014, from http://www.iea.cc/whats/index.html

International Nuclear Safety Advisory Group. (1986). *Summary report on the post-accident review meeting on the Chernobyl accident*. (Safety Series No. 75-INSAG-1). Vienna: International Atomic Energy Agency.

James, N., & Patrick, J. (2004). The Role of Situation Awareness in Sport. In S. Banbury & S. b. Tremblay (Eds.), *A cognitive approach to situation awareness: theory and application* (pp. 297-316). Burlington, VT: Ashgate.

Jin, H., & Wang, L. (2009). *Applying Situation Awareness to Human-Machine Interface Design of Aviation.* Paper presented at the Second International Symposium on Knowledge Acquisition and Modeling.

Johnson, C. (2004). *Final Report: Review of the BFU Überlingen Accident Report*. Contract C/1.369/HQ/SS/04. Eurocontrol.

Johnsen, S. O., & Veen, M. (2013). Risk assessment and resilience of critical communication infrastructure in railways. *Cognition, Technology & Work*, *15*(1), 95-107.

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space and Environmental Medicine, 67*(6), 507-512.

Jones, R. M., Laird, J. E., Nielsen, P. E., Coulter, K. J., Kenny, P., & Koss, F. V. (1999). Automated intelligent pilots for combat flight simulation. *AI Magazine, 20*(1), 27-41.

Jones, T., & Evans, D. (2000). Conducting a systematic review. *Australian Critical Care : Official Journal of the Confederation of Australian Critical Care Nurses, 13*(2), 66-71.

Jørgensen, K. M., & Ebrary. (2007). *Power without glory: a genealogy of a management decision*. Portland, OR: Copenhagen Business School.

Joseph, J. (2013). Resilience as embedded neoliberalism: A governmentality approach. *Resilience*, *1*(1), 38-52.

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics In Manufacturing, 10*(4), 409-430.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel Iii, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics, 36*(5), 447-462.

Kaempf, G. L., Klein, G., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human factors, 38*(2), 220-231.

Kärreman, D., Alvesson, M., Ekonomihögskolan vid Lunds universitet, E. H. L., Lunds, u., Lund University School of, E., Management, L., . . . Department of Business, A. (2009). Resisting resistance: counter-resistance, consent and compliance in a consultancy firm. *Human Relations, 62*(8), 1115-1144.

Kass, S. J., Cole, K. S., & Stanny, C. J. (2007). Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 10*(4), 321-329.

Keller, J. (2013). Karl Lilgert On Trial For BC Ferry Queen Of The North Sinking, *The Canadian Press*. Retrieved from http://www.huffingtonpost.ca/2013/01/17/karl-lilgert-bc-ferry-queen-of-the-north-sinking_n_2493214.html

Keller, J. (2014). Karl Lilgert's guilty verdict in fatal B.C. ferry sinking upheld by Appeal Court, *The Vancouver Sun*. Retrieved from http://www.vancouversun.com/news/Karl+Lilgert+guilty+verdict+fatal+ferry+sinking+upheld+Appeal+Court/10649178/story.html

Kendall, G., & Wickham, G. (2003). *Using Foucault's methods*. London: SAGE.

Khatwa, R., & Roelen, A. L. C. (1996). An Analysis of Controlled-flight-into-terrain (CFIT) Accidents of Commercial Operators, 1988 Through 1994. *Flight Safety Digest, 15,* 1-19.

Knight, R. G., Harnett, M., & Titov, N. (2005). The effects of traumatic brain injury on the predicted and actual performance of a test of prospective remembering. *Brain Injury, 19*(1), 19-27.

Knights, D., & McCabe, D. (2003). Governing through Teamwork: Reconstituting Subjectivity in a Call Centre. *JOURNAL OF MANAGEMENT STUDIES, 40*(7), 1587-1619.

Kolar, K. (2011). Resilience: Revisiting the concept and its utility for social research. *International Journal of Mental Health and Addiction*, *9*(4), 421-433.

Kokar, M. M., Matheus, C. J., & Baclawski, K. (2009). Ontology-based situation awareness. *Information Fusion, 10*(1), 83-98.

Kologlugil, S. (2010). Michel Foucault's archaeology of knowledge and economic discourse. *Erasmus Journal for Philosophy and Economics, 3*(2), 1-25.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction, 18*(1-2), 13-49.

Kuhn, T. S. (1996). *The structure of scientific revolution*. Chicago: University of Chicago Press.

Kumar, V., Rus, D., & Singh, S. (2004). Robot and sensor networks for first responders. *IEEE Pervasive Computing, 3*(4), 24-33.

LaPorte, T. R., & Consolini, P. M. (1991). Working in Practice but Not in Theory: Theoretical Challenges of "High-Reliability Organizations". *Journal of Public Administration Research and Theory: J-PART, 1*(1), 19-48.

Le Coze, J.-C. (2013). New models for new times. An anti-dualist move. *Safety Science Safety Science, 59*(1), 200-218.

Le Coze, J.-C., Pettersen, K., & Reiman, T. (2014). The foundations of safety science. *Safety Science, 67*(0), 1-5.

Lee, J. D., Cassano-Pinché, A., & Vicente, K. J. (2005). Bibliometric analysis of Human Factors (1970-2000): a quantitative description of scientific impact. *Human factors, 47*(4), 753-766.

Leveson, N. (2002). *A New Approach To System Safety Engineering*

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science, 42*(4), 237-270.

Leveson, N. (2012). Engineering a safer world systems thinking applied to safety. from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=421818

Leveson, N., Dulac, N., Marais, K., & Carroll, J. (2009). Moving Beyond Normal Accidents and High Reliability Organizations: A Systems Approach to Safety in Complex Systems. *Organization Studies, 30*(2-3), 227-249.

Lévi-Strauss, C. (1966). *The savage mind*. Chicago: University of Chicago Press.

Lindeman, R. W., Yanagida, Y., Noma, H., & Hosaka, K. (2006). Wearable vibrotactile systems for virtual contact and information display. *Virtual Reality, 9*(2-3), 203-213.

Lofquist, E.A. (2010). The art of measuring nothing: the paradox of measuring safety in a changing civil aviation industry using traditional safety metrics. *Safety Science, 48*, 1520-1529.

Loimer, H., & Guarnieri, M. (1996). Accidents and acts of God: a history of the terms. *American Journal of Public Health, 86*(1), 101-107.

Lu, C.S., & Tsai, C.L. (2010). The effect of safety climate on seafarers' safety behaviors in container shipping. *Accident Analysis and Prevention, 42*(6), 1999-2006.

Lu, C. S., & Yang, C.S. (2011). Safety climate and safety behavior in the passenger ferry context. *Accident Analysis and Prevention, 43*(1), 329-341.

Lundberg, J., & Rankin, A. (2013). Resilience and vulnerability of small flexible crisis response teams: Implications for training and preparation. *Cognition, Technology & Work.*

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics, 35*(10), 939-953.

Ma, R., & Kaber, D. B. (2007). Situation awareness and driving performance in a simulated navigation task. *Ergonomics, 50*(8), 1351-1364.

MacEachen, E. (2000). The mundane administration of worker bodies: From welfarism to neoliberalism. *Health, Risk & Society, 2*(3), 315-327.

Martin, J. (1992). *Cultures in Organizations: Three Perspectives*. Oxford University Press, New York.

McDonald, F. (2008). The criminalisation of medical mistakes in Canada: a review. *Health law journal, 16*, 1-25.

McKinlay, A., & Starkey, K. (1998). *Foucault, management and organization theory: from panopticon to technologies of self*. London: Sage Publications.

Meister, D. (1999). *The history of human factors and ergonomics*. Mahwah, N.J: Lawrence Erlbaum Associates.

Mengolini, A., & Debarberis, L. (2012). Lessons learnt from a crisis event: How to foster a sound safety culture. *Safety Science, 50*, 1415-1421.

Michaelides-Mateou, S., & Mateou, A. (2010). *Flying in the face of criminalization : the safety implications of prosecuting aviation professionals for accidents*. Burlington, VT: Ashgate.

Miller, A., & Xiao, Y. (2007). Multi-level strategies to achieve resilience for an organisation operating at capacity: A case study at a trauma centre. *Cognition, Technology & Work*, *9*(2), 51-66.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human factors, 49*(1), 57-75.

Miller, K. K., Riley, W., Davis, S., & Hansen, H. E. (2008). In situ simulation: A method of experiential learning to promote safety and team behavior. *Journal of Perinatal and Neonatal Nursing, 22*(2), 105-113.

Mills, S. (1997). *Discourse*. New York: Routledge.

Mills, S. (2003). *Michel Foucault*. New York: Routledge.

Ministry of Transport. (1968). *South African Airways Boeing 707-344C Aircraft, ZS-EUW, "Pretoria", at Windhoek on 20th April, 1968.*: Ministry of Transport.

Mishra, A., Catchpole, K., Dale, T., & McCulloch, P. (2008). The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surgical Endoscopy and Other Interventional Techniques, 22*(1), 68-73.

Moorkamp, M., Kramer, E., van Gulijk, C., & Ale, B. (2014). Safety management theory and the expeditionary organization: A critical theoretical reflection. *Safety Science*, *69*, 71-81.

Moray, N. (1994). "De Maximis Non Curat Lex" or how Context Reduces Science to Art in the Practice of Human Factors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 38*(9), 526-530.

Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science, 1*(4), 354-365.

Morel, G., Amalberti, R., & Chauvin, C. (2008). Articulating the differences between safety and resilience: The decision-making process of professional sea-fishing skippers.

Human Factors: The Journal of the Human Factors and Ergonomics Society, *50*(1), 1-16.

Morel, G., Amalberti, R., & Chauvin, C. (2009). How good micro/macro ergonomics may improve resilience, but not necessarily safety. *Safety Science*, *47*(2), 285-294.

Morris, B. T., & Trivedi, M. M. (2008a). Learning, modeling, and classification of vehicle track patterns from live video. *IEEE Transactions on Intelligent Transportation Systems, 9*(3), 425-437.

Morris, B. T., & Trivedi, M. M. (2008b). A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology, 18*(8), 1114-1127.

Moulton, C. A. E., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007). Slowing down when you should: A new model of expert judgment. *Academic Medicine, 82*(10 SUPPL.), S109-S116.

Murphy, R. R. (2004). Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 34*(2), 138-153.

Murray, S. J., Holmes, D., Perron, A., & Rail, G. (2007). No exit? Intellectual integrity under the regime of ‘evidence’ and ‘best-practices&rsquo. *Journal of Evaluation in Clinical Practice, 13*(4), 512-516.

Murray, S. J., Holmes, D., & Rail, G. (2008). On the constitution and status of `evidence' in the health sciences. *Journal of Research in Nursing, 13*(4), 272-280.

NAHQ (National Association for Healthcare Quality). (2012). *Call to Action: Safeguarding the Integrity of Healthcare Quality and Safety Systems*. Available at: www.nahq.org.

Nasar, J., Hecht, P., & Wener, R. (2008). Mobile telephones, distracted attention, and pedestrian safety. *Accident Analysis and Prevention, 40*(1), 69-75.

National Transportation Safety Board. (1974). *Aircraft Accident Report, Delta Air Lines, Inc., DC-9-31, N975NE, Boston, Maasachusetts, July 31, 1973*. (NTSB-AAR-74-3). Washington, D.C.: NTSB.

National Transportation Safety Board. (1976). *Aircraft Accident Report Alaska Airlines, Inc., B-727-81, N124AS, Ketchikan Internation Airport, Ketchikan, Alaska, April 5, 1976*. (NTSB-AAR-76-20). Washington, D.C.: NTSB.

Neisser, U. (1976). *Cognition and reality: principles and implications of cognitive psychology*. San Francisco: W.H. Freeman.

Nemeth, C., Wears, R. L., Patel, S., Rosen, G., & Cook, R. (2011). Resilience is not control: Healthcare, crisis management, and ICT. *Cognition, Technology & Work*, *13*(3), 189-202.

Nemeth, C. P., Nunnally, M., OConnor, M. F., Brandwijk, M., Kowalsky, J., & Cook, R. I. (2007). Regularly irregular: How groups reconcile cross-cutting agendas and demand in healthcare. *Cognition, Technology & Work*, *9*(3), 139-148.

Nixon, A., & Power, C. (2007). Towards a framework for establishing rigour in a discourse analysis of midwifery professionalisation. *Nursing inquiry, 14*(1), 71-79.

Norros, L., Liinasuo, M., & Savioja, P. (2014). Operators' orientations to procedure guidance in NPP process control. *Cognition, Technology & Work*, *16*(4), 487-499.

NPSA (National Patient Safety Agency). (2004). *Seven steps to patient safety: The full reference guide*. Second Print. Available at: www.nrls.npsa.nhs.uk

NRC (Nuclear Regulatory Commission). (2011). Final Safety Culture Policy Statement. *Federal Register, 76*(144), 34773-34778.

O'Farrell, C. (2006). Foucault and Postmodernism. *Sydney Papers, The, 18*(3/4), 182-194.

Owen, C., Healey, . N., & Benn, J. (2013). Widening the scope of human factors safety assessment for decommissioning. *Cognition, Technology & Work*, *15*(1), 59-66.

Page, S. E. (2008). Uncertainty, difficulty, and complexity. *Journal of Theoretical Politics*, *20*(2), 115.

Paltrinieri, N., Øien, K., & Cozzani, V. (2012). Assessment and comparison of two early warning indicator methods in the perspective of prevention of atypical accident scenarios. *Reliability Engineering & System Safety*, *108*, 21-31.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology, 21*(2), 270-297.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160.

Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human factors, 50*(3), 511-520.

Parker, I. (1992). *Discourse dynamics: critical analysis for social and individual psychology*. New York: Routledge.

Pasman, J., Knegtering, B., & Rogers, J. (2013). A holistic approach to control process safety risks: Possible ways forward. *Reliability Engineering & System Safety*, *117*, 21-29.

Paté-Cornell, M.E. (1993). Learning from the Piper Alpha accident: analysis of technical and organisational factors. *Risk Analysis, 13*, 215-232.

Patrick, J., & Morgan, P. L. (2010). Approaches to understanding, analysing and developing situation awareness. *Theoretical Issues in Ergonomics Science, 11*, 41-57.

Patterson, E. S., Woods, D. D., Cook, R. I., & Render, M. L. (2007). Collaborative cross-checking to enhance resilience. *Cognition, Technology & Work*, *9*(3), 155-162.

Patton, P. (1979). Of power and prisons. In M. Morris & P. Patton (Eds.), *Michel Foucault: Power/Truth/Strategy* (pp. 35-57). Sydney: Feral Publications.

Perrow, C. (1981). Normal Accident at Three Mile Island. *Society, 18*(5), 17-26.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies.* New York: Basic Books.

Perrow, C. (1999). *Normal accidents*. Princeton (NJ): Princeton university press.

Peters, T.J., & Waterman, R.H. (1982). *In Search of Excellence*. New York, NY: Harper & Row.

Petrovskaya, A., & Thrun, S. (2009). Model based vehicle detection and tracking for autonomous urban driving. *Autonomous Robots, 26*(2-3), 123-139.

Petticrew, M. (2001). Systematic reviews from astronomy to zoology: myths and misconceptions. *BMJ (Clinical research ed.), 322*(7278), 98-101.

Pickering, C. M., & Byrne, J. A. (2013). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers.

Pidgeon, N. (1998). Safety culture: key theoretical issues. *Work and Stress, 12*, 202-216.

Pidgeon, N., & O'Leary, M. (2000). Man-made disasters: why technology and organizations (sometimes) fail. *Safety Science, 34*(1), 15-30.

Powers, P. (2007). The Philosophical Foundations of Foucaultian Discourse Analysis. *Critical Approaches to Discourse Analysis across Disciplines, 1*(2), 18-34.

Ramos do Ó, J. (2005). Notas sobre Foucault e a governamentalidade (pp. 15-39). In: L.S., Falcão; P. Souza. (eds.) *Michael Foucault: perspectivas*. Rio de Janiero: Achiamé.

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Studies International Journal of Human-Computer Studies, 45*(5), 579-597.

Rankin, A., Dahlbäck, N., & Lundberg, J. (2013). A case study of factor influencing role improvisation in crisis response teams. *Cognition, Technology & Work*, *15*(1), 79-93.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, 13*, 257-266.

Rasmussen, J. (1990a). Human Error and the Problem of Causality in Analysis of Accidents. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 327*(1241), 449-462.

Rasmussen, J. (1990b). The role of error in organizing behavior. *Ergonomics, 33*(10/11), 1185-1199.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science, 27*(2), 183-213.

Rasmussen, J. (2000). Human factors in a dynamic information society: where are we heading? *Ergonomics, 43*(7), 869-879.

Rasmussen, J. (2010). Enabling selves to conduct themselves safely: Safety committee discourse as governmentality in practice. *Human Relations, 64*(3), 459-478.

Rasmussen, J. (2013). Governing the workplace or the worker? Evolving dilemmas in chemical professionals' discourse on occupational health and safety. *Discourse and Communication, 7*(1), 75-94.

Re, A., & Macchi, L. (2010). From cognitive reliability to competence? An evolving approach to human factors and safety. *Cognition, Technology & Work*, *12*(2), 79-85.

Reader, T., Flin, R., Lauche, K., & Cuthbertson, B. H. (2006). Non-technical skills in the intensive care unit. *British Journal of Anaesthesia, 96*(5), 551-559.

Reason, J. T. (1990). *Human error*. Cambridge [England]; New York: Cambridge University Press.

Reason, J. T. (1997). *Managing the risks of organizational accidents*. Aldershot, England; Brookfield, Vt., USA: Ashgate.

Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human factors, 42*(3), 349-366.

Richter, A., & Koch, C. (2004). Integration, differentiation and ambiguity in safety cultures. *Safety Science, 42*, 703-722.

Robertson, M. M., & Endsley, M. R. (1997). *Development of situation awareness training program for aviation maintenance.* Paper presented at the Annual Meeting of the Human Factors and Ergonomics Society.

Rochlin, G. I. (1999). Safe operation as a social construct. *Ergonomics, 42*(11), 1549-1560.

Rochlin, G. I., La Porte, T. R., & Roberts, K. H. (1987). The self-designing high-reliability organization: Aircraft carrier flight operations at sea. *Naval War College Review*, 76-90.

Rollenhagen, C. (2010). Can focus on safety culture become an excuse for not rethinking design of technology? *Safety Science, 48*, 268-278.

Roscoe, S. N. (1997). The Adolescence of Engineering Psychology. *Human Factors History Monograph Series, 1*.

Rose, N. S. (1996). *Inventing our selves: psychology, power, and personhood*. Cambridge, England: Cambridge University Press.

Rose, N. S. (1996). Power and Subjectivity: Critical History and Psychology. In C. F. Graumann & K. J. Gergen (Eds.), *Historical dimensions of psychological discourse* (pp. 103-124). New York: Cambridge University Press.

Ross, J., Anderson, . E., Kodate, N., Thompson, K., Cox, A., & Malik, R. (2012). Inpatient diabetes care: Complexity, resilience and quality of care. *Cognition, Technology & Work.*

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators and Virtual Environments, 11*(4), 335-351.

Rupert, A. H. (2000). An instrumentation solution for reducing spatial disorientation mishaps: A more 'natural' approach to maintaining spatial orientation. *IEEE Engineering in Medicine and Biology Magazine, 19*(2), 71-80.

Sagan, S.D. (1993). The Limits of Safety: Organisations, accidents, and nuclear weapons. Princeton, NJ: Princeton University Press.

Salas, E., & Dietz, A. S. (2011). *Situational awareness*. Burlington, VT: Ashgate.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human factors, 37*(1), 123-136.

Saleh, J. H., & Marais, K. (2006). Highlights from the early (and pre-) history of reliability engineering. *Reliability Engineering and System Safety, 91*(2), 249-256.

Salmon, P. M., Stanton, N. A., Walker, G. H., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied ergonomics, 37*(2), 225-238.

Salmon, P. M., Stanton, N. A., Jenkins, D. P., Walker, G. H., Young, M. S., & Aujla, A. (2007). What Really Is Going on? Review, Critique and Extension of Situation Awareness

Theory. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (Vol. 4562, pp. 407-416). Berlin, Heidelberg: Springer.

Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., & Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science, 9*(4), 297-323.

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., Ladva, D., Rafferty, L., & Young, M. S. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics, 39*(3), 490-500.

Salmon, P. M., Walker, G. H., & Stanton, N. A. (2015). Broken components versus broken systems: Why it is systems not people that lose situation awareness. *Cognition, Technology and Work, 17*(2), 179-183.

Sarter, N. B., & Woods, D. D. (1991). Situation Awareness: A Critical But Ill-Defined Phenomenon. *The International Journal of Aviation Psychology, 1*(1), 45-57.

Sarter, N. B., & Woods, D. D. (1995). How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human factors, 37*(1), 5.

Saurin, T. A., & Carim Júnior, G. C. (2011). Evaluation and improvement of a method for assessing HSMS from the resilience engineering perspective: A case study of an electricity distributor. *Safety Science*, *49*(2), 355-368.

Saussure, F. d. (1974). *Course in General Linguistics*. London: Fontana.

Savioja, P., Norros, L., Salo, L., & Aaltonen, I. (2014). Identifying resilience in proceduralised accident management activity of NPP operating crews. *Safety Science*, *68*, 258-274.

Sawyer, R. K. (2002). A discourse on discourse: An archeological history of an intellectual concept. *CULTURAL STUDIES, 16*(3), 433-456.

Schein, E.H. (2004). *Organisational Culture an Leadership* (3rd Ed.). San Francisco: Jossey-Bass.

Schmidt, J. (2013). The empirical falsity of the human subject: New materialism, climate change and the shared critique of artifice. *Resilience*, *1*(3), 174-192.

Schmoch, U., & Schubert, T. (2008). Are international co-publications an indicator for quality of scientific research? *Scientometrics, 74*(3), 361-377.

Schraagen, J. M. (2013). To publish or not to publish: A systems analysis of longitudinal trends in publishing strategies of a human factors research organisation. *Theoretical Issues in Ergonomics Science*, *14*(5), 499-530.

Schulz, C. M., Endsley, M. R., Kochs, E. F., Gelb, A. W., & Wagner, K. J. (2013). Situation awareness in anesthesia: concept and research. *Anesthesiology, 118*(3), 729-742.

Sebok, A. (2000). Team performance in process control: Influences of interface design and staffing levels. *Ergonomics, 43*(8), 1210-1236.

Seeck, H., & Kantola, A. (2009). Organizational control: Restrictive or productive? *J Man Org Journal of Management & Organization, 15*(2), 241-257.

Selker, T., & Burleson, W. (2000). Context-aware design and interaction in computer systems. *IBM Systems Journal, 39*(3-4), 880-891.

Sengupta, R., Rezaei, S., Shladover, S. E., Cody, D., Dickey, S., & Krishnan, H. (2007). Cooperative collision warning systems: Concept definition and experimental implementation. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 11*(3), 143-155.

Sevdalis, N., Davis, R., Koutantji, M., Undre, S., Darzi, A., & Vincent, C. A. (2008). Reliability of a revised NOTECHS scale for use in surgical teams. *American Journal of Surgery, 196*(2), 184-190.

Sevdalis, N., Forrest, D., Undre, S., Darzi, A., & Vincent, C. (2008). Annoyances, disruptions, and interruptions in surgery: The Disruptions in Surgery Index (DiSI). *World Journal of Surgery, 32*(8), 1643-1650.

Sharma, O. P. (2012). Quality Indicators of Scientific Research. *Indian Journal of Microbiology, 52*(2), 305-306.

Sheen, M.J. (1987). *MV Herald of Free Enterprise Report of Court No. 8074 Formal Investigation*. London: HMSO.

Sheridan, . B. (2008). Risk, human error, and system resilience: Fundamental ideas. *Human Factors*, *50*(3), 418-426.

Sheth, A. (2009). Citizen sensing, social signals, and enriching human experience. *IEEE Internet Computing, 13*(4), 87-92.

Shirali, G., Mohammadfam, I., & Ebrahimipour, V. (2013). A new method for quantitative assessment of resilience engineering by PCA and NT approach: A case study in a process industry. *Reliability Engineering & System Safety*, *119*, 88-94.

Shrivastava, P. (1987). *Bhopal: Anatomy of a Crisis*. Cambridge, MA: Ballinger.

Shrivastava, S., Snopar, K., & Pazzaglia, F. (2009). Normal Accident Theory versus High Reliability Theory: A resolution and call for an open systems view of accidents. *Human Relations, 62*(9), 1357-1390.

Sibley, S.S. (2009). Taming Prometheus: Talk about safety and culture. *The Annual Review of Sociology, 35*, 341-369.

Simon, H. A. (1980). The behavioral and social sciences. *Science, 209*(4452), 72-78.

Singh, H., Petersen, L. A., & Thomas, E. J. (2006). Understanding diagnostic errors in medicine: A lesson from aviation. *Quality and Safety in Health Care, 15*(3), 159-164.

Skjerve, B., Kaarstad, M., Størseth, F., Wærø, I., & Grøtan, O. (2012). Planning for resilient collaboration at a new petroleum installation – A case study of a coaching approach. *Safety Science*, *50*(10), 1952-1959.

Smith, M.J., Cohen, H.H., & Cohen, A. (1978). Characteristics of successful safety programs. *Journal of Safety Research, 10*, 5-15.

Smith, K., & Hancock, P. A. (1995). Situation Awareness Is Adaptive, Externally Directed Consciousness. *Human factors, 37*(1), 137-148.

Snook, S. A. (2000). *Friendly fire: the accidental shootdown of U.S. Black Hawks over Northern Iraq*. Princeton, N.J.: Princeton University Press.

Sonnenwald, D. H., & Pierce, L. G. (2000). Information behavior in dynamic group work contexts: Interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing and Management, 36*(3), 461-479.

Sorensen, J.N. (2002). Safety Culture: A survey of the sate-of-the-art. *Reliability Engineering and System Safety, 76*, 189-2004.

Stanton, N. A. (2006). Situation(al) Awareness: Alternative Theories. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 891-896). Boca Raton, FL: Taylor & Francis.

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science, 39*(3), 189-204.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P. M., Hoyle, G., Walker, G. H., Young, M. S., Linsell, M., Dymott, R., & Green, D. (2006). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. *Ergonomics, 49*(12-13), 1288-1311.

Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics, 48*(10), 1294-1313.

Starbuck, W. H., & Farjoun, M. (2005). *Organization at the Limit: Lessons from the Columbia Disaster*. Oxford, UK: Blackwell Publishing.

Starbuck, W., & Milliken, F. (1988). Challenger: fine-tuning the odds until something breaks. *Journal of Management Studies, 25*, 319-340.

Starkey, L. J., & Maeda, S. (2009). Doctor as criminal: reporting of patient deaths to the police and criminal prosecution of healthcare providers in Japan. *BMC health services research, 10*(1), 53-53.

Steen, R., & Aven, T. (2011). A risk perspective suitable for resilience engineering. *Safety Science*, *49*(2), 292-297.

Suchman, L. A. (2007). *Human-machine reconfigurations: plans and situated actions*. Cambridge; New York: Cambridge University Press.

Supreme Court of British Columbia. (2013). *Regina v. Karl-Heinz Arthur Lilgert*. (2013 BCSC 1329). Vancouver.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. (1997). Information complexity - Mental workload and performance in combat aircraft. *Ergonomics, 40*(3), 362-380.

Swell, G., Wilkinson, B. (1992). "Someone to watch over me": Surveillance, discipline and the just-in-time labour process. *Sociology, 26*, 271-289.

Taylor, F. W. (1967). *The principles of scientific management*. New York: Norton.

Tenney, Y. J., & Pew, R. W. (2006). Situation Awareness Catches On: What? So What? Now What? *Reviews of Human Factors and Ergonomics, 2*(1), 1-1.

Tharaldsen, J., & Haukelid, K. (2009). Culture and behavioural perspectives on safety – towards a balanced approach. *Journal of Risk Research, 12*(3-4), 375-388.

Thomas, G. (2007). A Crime Against Safety. *Air Transport World, 44*(1), 57.

Titscher, S., Meyer, M., Wodak, R., & Vetter, E. (2000). *Methods of Text and Discourse Analysis*. Sage.

Toet, A., & Franken, E. M. (2003). Perceptual evaluation of different image fusion schemes. *Displays, 24*(1), 25-37.

Toet, A., Ijspeert, J. K., Waxman, A. M., & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness. *Displays, 18*(2), 85-95.

Topp, W. (2000). Knowledge System Diagnostics: Applying Foucault's Archaeological Framework to Organizations. *SYSTEMS RESEARCH AND BEHAVIORAL SCIENCE, 17*, 365-374.

Transportation Safety Board of Canada. (2003). *In-flight fire leading to collision with water : Swissair Transport Limited, McDonnell Douglas MD-11 HB-1WF, Peggy's Cove, Nova Scotia 5nm SW, 2 September 1998*. [Aylmer, Quebec]: Transportation Safety Board of Canada.

Transportation Safety Board of Canada. (2008). *Striking and Subsequent Sinking, Passenger and Vehicle Ferry Queen of the North, Gil Island, Wright Sound, British Columbia, 22 March 2006*. (M06W0052). TSB.

Turner, K. (1989). Safety, Discipline and the Manager: Building a `Higher Class of Men'. *Sociology, 23*(4), 611-628.

Turner, N., & Gray, G.C. (2009). Socially constructing safety. *Human Relations, 62*(9), 1259-1266.

Tveiten, C. K., Albrechtsen, E., Wærø, I., & Wahl, A. M. (2012). Building resilience into emergency management. *Safety Science*, *50*(10), 1960-1966.

Tweedale, J., Ichalkaranje, N., Sioutis, C., Jarvis, B., Consoli, A., & Phillips-Wren, G. (2007). Innovations in multi-agent systems. *Journal of Network and Computer Applications, 30*(3), 1089-1115.

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics, 50*(8), 1235-1249.

Ungar, M. (2005). *Handbook for working with children and youth: Pathways to resilience across cultures and contexts.* Sage.

van Westrenen, F. (2014). Modelling arrival control in a vessel traffic management system. *Cognition, Technology & Work*, *16*(4), 501-508.

Vaughan, D. (1996). *The Challenger launch decision: risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press.

Vaughan, D. (1999). The dark side of organizations: mistake, misconduct and disaster. *Annual Review of Sociology, 25*, 271-305.

Vaughan, D. (2004). Theorizing disaster: analogy, historical ethnography, and the Challenger accident. *Ethnography, 5*, 313-345.

Vaughan, D. (2006). NASA revisited: theory, analogy, and public sociology. *American Journal of Sociology, 112*, 353-393.

Vicente, K. J. (1999). Cognitive work analysis: Toward safe, productive, and healthy computer-based work. Mahwah, NJ: Lawrence Erlbaum Associates.

Vinodkumar, M.N., & Bhasi, M. (2010). Safety management practices and safety behaviour: Assessing the mediating role of safety knowledge and motivation. *Accident Analysis and Prevention, 42*(6), 2082-2093.

von Bertalanffy, L. (1950). An outline of general system theory. *British Journal for the Philosophy of Science*, *1*, 139-164.

von Bertalanffy, L. (1969). *General system theory; foundations, development, applications*. New York: G. Braziller.

Waitt, G. (2005). Doing Discourse Analysis. In I. Hay (Ed.), *Qualitative research methods in human geography* (pp. 163-191). South Melbourne, Vic: Oxford University Press.

Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, adaptability and transformability in social--ecological systems. *Ecology and Society*, *9*(2), 5.

Walker, J., & Cooper, M. (2011). Genealogies of resilience: From systems ecology to the political economy of crisis adaptation. *Security Dialogue*, *42*(2), 143-160.

Wang, W., Mao, Y., Jin, J., Wang, X., Guo, H., Ren, X., & Ikeuchi, K. (2011). Driver's various information process and multi-ruled decision-making mechanism: A fundamental of intelligent driving shaping model. *International Journal of Computational Intelligence Systems, 4*(3), 297-305.

Watson, J. B. (1913). Psychology as the Behaviorist Views It. *Psychological Review, 20*(2), 158-177.

Weaver, S. J., Lyons, R., Diazgranados, D., Rosen, M. A., Salas, E., Oglesby, J., . . . King, H. B. (2010). The anatomy of health care team training and the state of practice: A critical review. *Academic Medicine, 85*(11), 1746-1760.

Weick, K.E. (1990). The vulnerable system: an analysis of the Tenerife air disaster. *Journal of Management, 16*, 571-593.

Weick, K. E. (1993). The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Administrative Science Quarterly, 38*(4), 628-652.

Weick, K. E. (2001). *Making sense of the organization*. Oxford, UK; Malden, MA: Blackwell Publishers.

Weick, K.E. (2004). Normal Accident Theory as Frame, Link, and Provocation. *Organization & Environment, 17*(1), 27-31.

Westrum, R. (2006). A typology of resilience situations. In E. Hollnagel, D. Woods, & N. Leveson (Eds.), *Resilience engineering - concepts and precepts* (pp. 43-53). Aldershot: Ashgate Publishing Company.

Wetherell, M., Yates, S., & Taylor, S. (2001). *Discourse theory and practice: a reader*. Thousand Oaks, [Calif.]: SAGE.

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

Wickens, C. D. (2008). Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human factors, 50*, 397-403.

Wiegmann, D.A. et al. (2004). Safety Culture: An Integrative Review. *The International Journal of Aviation Psychology, 14*(2), 117-134.

Wiegmann, D.A., & Shappell, S.A. (2003). A human error approach to aviation accident analysis: The Human Factors Analysis and Classification System. Aldershot, UK: Ashgate.

Willig, C. (2008). *Introducing qualitative research in psychology: adventures in theory and method*. Maidenhead, English: McGraw Hill/Open University Press.

Willmott, H. (1993). "Strength is ignorance, slavery is freedom": managing culture in modern organisations. *Journal of Management Studies, 30*, 515-552.

Wilson, J. R. (2000). Fundamentals of ergonomics in theory and practice. *Applied ergonomics, 31*(6), 557-567.

Wilson, J. R. (2014). Fundamentals of systems ergonomics/human factors. *Applied ergonomics, 45*(1), 5-13.

Wilson, J. R., Ryan, B., Schock, A., Ferreira, P., Smith, S., & Pitsopoulos, J. (2009). Understanding safety and production risks in rail engineering planning and protection. *Ergonomics*, *52*(7),

Winsen, R. v., Henriqson, E., Schuler, B., & Dekker, S. W. A. (2015). Situation awareness: some conditions of possibility. *Theoretical Issues in Ergonomics Science, 16*, 53-68.

Wolbers, T., Hegarty, M., Büchel, C., & Loomis, J. M. (2008). Spatial updating: How the brain keeps track of changing object locations during observer motion. *Nature Neuroscience, 11*(10), 1223-1230.

Won, J. S., & Langari, R. (2005). Intelligent energy management agent for a parallel hybrid vehicle - Part II: Torque distribution, charge sustenance strategies, and performance results. *IEEE Transactions on Vehicular Technology, 54*(3), 935-953.

Woods, D. D. (2006). Resilience engineering: Redefining the culture of safety and risk management. *Hum Factors Ergonom Soc Bull*, *49*, 1-3.

Woods, D. D., & Branlat, M. (2011). Basic petterns in how adaptive systems fail. In E. Hollnagel, J. Pariès, D. D. Woods, & J. Wrethall (Eds.), *Resilience engineering in practice - A guidebook* (pp. 127-143). Farnham Surrey: Ashgate Publishing Limited.

Woods, D. D., & Cook, R. I. (2002). Nine Steps to Move Forward from Error. *Cognition, Technology & Work, 4*(2), 137-144.

Woods, D. D., Dekker, S. W. A., Cook, R., & Sarter, N. B. (2010). *Behind Human Error*. Farnham, England: Ashgate.

Woods, D. D., Patterson, E. S., & Roth, E. M. (2002). Can We Ever Escape from Data Overload? A Cognitive Systems Diagnosis. *Cognition, Technology & Work, 4*(1), 22-36.

Woods, D. D., Sarter, N. B., Cook, R. I., & Johannesen, L. J. (1994). *Behind human error: cognitive systems, computers, and hindsight*.

Woods, D. D., Schenk, J., & Allen, T. T. (2009). An initial comparison of selected models of system resilience. In C. P. Nemeth, E. Hollnagel, & S. W. A. Dekker (Eds.), *Resilience engineering perspectives* (Vol. 2). Farnham: Ashgate Publishing Limited.

Woods, D. D., & Wrethall, J. (2008). Stress-strain plots as a basis for assessing system resilience. In E. Hollnagel, C. P. Nemeth, & S. W. A. Dekker (Eds.), *Resilience engineering perspecties* (Vol. 1, pp. 143-158). Aldershot: Ashgate Publishing Limited.

Wright, M. C., Taekman, J. M., & Endsley, M. R. (2004). Objective measures of situation awareness in a simulated medical environment. *Quality and Safety in Health Care, 13*(SUPPL. 1), i65-i71.

Xu, Q. (2000). On the Way to Knowledge: Making a Discourse at Quality. *Organization, 7*(3), 427-453.

Yee, B., Naik, V. N., Joo, H. S., Savoldelli, G. L., Chung, D. Y., Houston, P. L., . . . Hamstra, S. J. (2005). Nontechnical skills in anesthesia crisis management with repeated exposure to simulation-based education. *Anesthesiology, 103*(2), 241-248.

Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G., & Paterson-Brown, S. (2008). Surgeons' non-technical skills in the operating room: Reliability testing of the NOTSS behavior rating system. *World Journal of Surgery, 32*(4), 548-556.

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006). Development of a rating system for surgeons' non-technical skills. *Medical Education, 40*(11), 1098-1104.

Zeeman, L., Poggenpoel, M., Myburgh, C. P. H., & Van der Linde, N. (2002). An introduction to a postmodern approach to educational research: Discourse analysis. *Education, 123*(1), 96.

Zhang, H., et al. (2002). Safety culture: A concept in chaos? In: *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: HFES.

Zhang, Y., Markham, P., Xia, T., Chen, L., Ye, Y., Wu, Z., & Liu, Y. (2010). Wide-area frequency monitoring network (FNET) architecture and applications. *IEEE Transactions on Smart Grid, 1*(2), 159-167.

Zieba, S., Polet, P., Vanderhaegen, F., & Debernard, S. (2010). Principles of adjustable autonomy: A framework for resilient human–machine cooperation. *Cognition, Technology & Work*, *12*(3), 193-203.

Zohar, D. (1980). Safety Climate in Industrial Organizations: theoretical and applied implications. *Journal of Applied Psychology, 65*, 96-102.

Zoller, H.M. (2003). Health on the line: Identity and disciplinary control in employee occupational health and safety discourses. *Journal of Applied Communication Research, 31*(2), 118-139.

Appendix A.

Human factors and the ethics of explaining failure

This appendix includes a book chapter from an edited collection. The copyright of this article has not been transferred to Ashgate, however, authors retain the right to include their manuscript in a thesis or dissertation that is not to be published commercially.

The bibliographic details of the co-authored paper, including all authors, are:

Winsen, R. van, & Dekker, S. W. A. (submitted). Human Factors and the ethics of explaining failure. In: S. Shorrock & C. Williams (Eds.). *Human Factors and Ergonomics in Practice.* Aldershot, UK: Ashgate Publishing Co.

As a the first-author of this paper I have had a leading role in pursuing this argument; I have reviewed the literature; I have engaged the co-authors in discussions about the arguments pursued in the chapter; I have designed and written the chapter; I have implemented the feedback from the co-author; I have written all the text that makes up the final version of the paper; I have critically revised numerous versions of the paper; I have implemented the feedback from the book's editors.

(Signed) ________________ (Date) 22-02-2015
Roel van Winsen

(Countersigned) ________________ (Date) ________
Sidney W. A. Dekker (Principal supervisor and co-author)

Human factors and the ethics of explaining failure

Introduction

Karl is in jail today. And human factors might have had something to do with it. Karl may not be the last one. Karl Lilgert was the officer in charge of the large passenger ferry 'Queen of the North' when it ran aground and sank. Ninety-nine passengers and crew survived the sinking of the ferry. Unfortunately, two passengers were never found and presumed to have drowned. Despite Karl facing poor weather, substandard navigational equipment, inadequate company policies, and a lack of staff on the bridge, he was essentially blamed for the failure to turn the ferry in time (Keller, 2013). Karl was criminally prosecuted, particularly as the 'causes and contributing factors' in the accident investigation report strongly suggested that various distractions contributed to his failure to order the required course change (Transportation Safety Board of Canada, 2008). Additionally, the investigation report concluded:

> "*The working environment on the bridge of the Queen of the North was less than formal, and the accepted principles of navigation safety were not consistently or rigorously applied. Unsafe navigation practices persisted which, in this occurrence, contributed to the loss of situational awareness by the bridge team*."
> (Transportation Safety Board of Canada, 2008, p. 6)

As the court held that, "[m]aintaining situational awareness at all times and in all circumstances is key to proper navigation" (Supreme Court of British Columbia, 2013), the leap wasn't large: the prosecution was able to successfully argue for the dereliction of his duty. Evidently, situational awareness was not maintained, otherwise the ferry would not have crashed and sunk. Karl was convicted for criminal negligence causing the deaths of the two passengers and sentenced to four years in prison. He is there at the time of writing this.

Human factors has embraced 'situation awareness' (SA) as a construct to aid our understanding of human decision making in complex dynamic systems and to help with the design of human-machine interfaces. However, now SA – and the loss of it – functions as a causal and normative construct to explain and judge human performance (at least in the case of Karl). This provides an interesting paradox, particularly as the primary concern of human factors is with the well-being of humans in interaction with technology, or better, systems. According to the International Ergonomics Association (2014), human factors and ergonomics is "*the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance*." As such, one of the discipline's central challenges lies in explaining and preventing accidents in our high-risk socio-technological systems. In this chapter we will focus on the ethical implications of the popular notion of 'human error' – and its subcategories, such as 'loss of SA' – as explanations for why sometimes socio-technological systems fail.

It is here – in human error explanations of accidents – that the tension between systems and individuals creates some paradoxes that need both practical and ethical consideration, as exemplified by the case of Karl. The basic tension is one between the parts and the whole. After outlining the historical imperative of human factors as a systems science, in this chapter, we discuss the tendency to default to 'human error', particularly in human factors explanations of accidents. This unstated – and possibly unrecognised – commitment to reductionism, that regards the human as a largely independent but problematic component of our complex socio-technological systems, has resulted in a number of concepts and practices in human factors whose effects run counter to the discipline's aim of optimising human well-being. As such, it is pertinent that we, as a human factors community, engage in ethical discussions about the goals, methods, and consequences of our discipline's propositions.

Human factors versus human error: systems and components

Historically, human performance – in terms of efficiency and safety – became of interest with the rise of the industrial society at the turn of the 19th century. Increasing numbers of miners, factory workers, and other operators were exposed to new machines, working conditions and risks. Matching humans to new technology was done through increasingly advanced selection procedures and training programmes, intended to better fit the human to the system. Few efforts were made to adjust the system to fit the human. However, with the rapid development of even more sophisticated technologies during the 20th century, it became increasingly necessary to adjust the task or tools to match the human, that is, to design equipment that takes into consideration the needs and limits of the human (Meister, 1999).

A seminal study that set the agenda for the scientific discipline of human factors was by the experimental psychologists Fitts and Jones (1947), who adapted their laboratory techniques to study the applied problem of 'pilot error' during WWII. The problem they faced was that pilots of one aircraft type frequently retracted the gear instead of the flaps after landing. This incident hardly ever occurred to pilots of other aircraft types. They noticed that the gear and flap controls could easily be confused: the nearly identical levers were located right next to each other in an obscure part of the cockpit. Conversely, the flap and gear controls in the cockpits of the control group of aircraft were not adjacent. Fitts and Jones realised that the gear-up-after-landing problems – the 'pilot errors' – were cockpit-design issues instead. Alphonse Chapanis, another human factors pioneer, fixed a small rubber wheel and a small wedge-shaped end respectively, to the landing gear lever and the flap lever, and this type of 'human error' almost completely disappeared. This insight, that human performance is to be understood as being systematically connected to the features of people's tools and tasks, effectively constitutes the birth of human factors.

Later, particularly in the wake of a number large organisational accidents in the late 1970's and 1980's, Fitts and Jones' principles for the design of tools and interfaces was extended to include the broader influence of organisational environments. Reason (1990), for example, proposed that the performance of those at the sharp (operational) end is shaped by latent conditions and upstream organisational factors at the blunt (managerial) end. Organisational pressures for production and efficiency slowly but systematically shape the decision-making and performance of those in the system (see Dekker's (2011) 'drift into failure'). Human performance and behaviour is connected to the operational and organisational environments people work in. By presuming that human performance can only be understood by seeing it as embedded in the larger system in which it takes place, this kind of human factors builds on a holistic – systemic – worldview that sees the system as an integrated whole rather than a collection of disconnected parts. Properties of the system cannot be understood by analysing the parts, but emerge from the interactions and relationships among the components. In terms of accident explanation and prevention, this means that rather than seeing accidents as resultant of (a linear succession of) failing components, accidents are seen as emergent phenomena that cannot be understood by identifying faulty components – such as the human in error.

That said, accidents are often still seen as the result of risk that was not managed well, either at the sharp operational end or the blunt end (whether management, design, construction, or maintenance). Whereas in the 1960's 'human error' was said to be the main cause of accidents in about 30% of all accidents, it is now commonly accepted that 70% to 90% of accidents can be attributed to 'human error' (Hollnagel, 2006a). This seems at odds with the human factors principle that human performance is systematically connected to the system in which it takes place. Instead of seeing an accident as the emergent outcome of a complex system, and which is thus destroyed when the phenomenon is reduced to isolated components, accident explanations seem to rely on a reductionist search for the 'broken part' that caused the accident. The belief that the behaviour of the whole can be understood by analysing the properties of its parts is central to Descartes's method of analytic reductionism. The hegemony of analytic reductionism in our Western scientific approach also shines through in human factors practices, particularly those concerned with accident explanation and prevention.

One possible explanation for this trend towards human error explanations of accidents may lie in the inherent contradiction between human factors' pragmatic orientation and the systems approach it tries to build upon. As a practical discipline, human factors aims to provide (design-) recommendations to improve the interaction of humans and systems. While (reductionist) science often isolates its objects from their context – to study them in controllable experimental settings – in order to be able to provide causal explanations, systems approaches rely on a more intuitive and holistic approach that does not yield the same 'practical' truths (in terms of cause and effect). Systems explanations point away from single-cause explanations, and seek multiple diverging narratives that enlarge the scope of analysis to understand emergent phenomena. As small changes may have large effects in systems theory, at best, it hopes to find coexistence and correlation – as opposed to causation – in the interactions and relations that make up our socio-technological systems. Consequently, after fixing the obvious, manifest problems in the system, it becomes increasingly difficult for human factors as a systems science to provide 'simple' and practical solutions: where do we intervene or re-design systems when everything in the system is interconnected and interdependent? When we cannot locate a cause for the accident, how can we fix the system so that things do not go wrong again? Also, very important nowadays, from a systems perspective it may be difficult (if not impossible) to understand who is responsible for any harm.

In understanding why accident investigation reports often set out to explore multiple diverging narratives but then 'suddenly' converge in single localised explanations, Galison (2000, p. 40) suggested:

> "*… if there is no seed, if the bramble of cause, agency, and procedure does not issue from a fault nucleus, but is rather unstably perched between scales, between the human and non-human, and between protocol and judgment, then the world is a more disordered and dangerous place. These reports, and much of the history we write, struggle, incompletely and unstably, to hold that nightmare at bay*."

This is where reductionist explanations are useful: Finding a broken component – 'human error' – provides us with an endpoint of an accident investigation, a root cause. It gives us something to fix.

Human factors and ethics: the implications of some human error practices

Whereas systems thinking sees the world as a complex web of interactions and relationships, human error is still regarded as a practical target for intervention. This comes with a price – there might be an ethical cost to our reductionist approach. Ethics is the branch of philosophy that is concerned with what the right thing is to do, what is good for individuals and society. Of course, explanations are always simplifications of reality; otherwise our theories and models would be as complex as the phenomena they aim to understand and would yield no explanatory or predictive value. The question is – and here it becomes an ethical issue – what knowledge or solutions do we want our theories or explanations to produce? In the wake of an accident, what countermeasures do we hope to launch, what do we miss by relying on these explanations, and who wins and who loses? Are our practices serving the well-being of society or the operators that human factors claims to have in its best interest? By unearthing the reductionist assumptions underlying some human factors practices, in this section we ask ethical questions of three ways of dealing with human error – error taxonomies, automation, and proceduralisation.

Error taxonomies & categories

In trying to reconcile the tension between practical solutions and its systems ideas, human factors has proposed a great number of performance and error taxonomies intended to provide researchers or accident investigators with a 'deeper insight' into understanding socio-technological failures. We have developed classifications based on different types of work, various cognitive failures, error types corresponding to performance modes (skill-based slips and lapses, rule-based mistakes, and knowledge-based mistakes), and, more recently, taxonomies that are supposed to support the practical management of safety (such as threat and error management [TEM] and human factors analysis and classification system [HFACS]). In addition to these more extensive classification systems, human factors has also proposed and popularized a number of (single) categories that aim to provide specificity to errors: SA, complacency, automation surprise, breakdown of CRM, etc. All these taxonomies and constructs are developed with the best of intentions – that is, in line with the human factors systems philosophy – as tools to help to better understand where in the system the failure is located.

One issue with these taxonomies, however, is their fragmented focus on components of the system, most prominently the human. Where human factors is supposed to be about the interaction of humans and technology or systems, very few human factors theories or classification tools take these interactions as its main subject – cognitive systems engineering is the only evident exception, as it explicitly takes the joint cognitive system as its unit of analysis. Otherwise, these classifications provide slightly more specific labels for component failures that often remove the operational, organisational, and social context in which the behaviour took place. This (retrospective) labelling of human performance into more general error types removes context and so hampers our understanding of the situation that the people involved in the accident were facing.

As a result, many of the taxonomies and categories we use do not provide a satisfactory explanation for why the person did what he or she did; for why a decision or action made sense at the time. The 'local rationality principle' dictates that people do what is rational given the knowledge, goals, and resources that they have available at the moment of the decision or action. Human error classifications and explanations do not take the perspective of the people who were facing a particular situation, but rather judge their behaviour based on all the information and resources we have available in hindsight. Not only do these 'judgments' – based on information that we now, in hindsight, know to be important – provide normative standards for behaviour (e.g. to always be 'situationally aware'), the language that comes with them often allows for naïve allocations of blame. Also, by disregarding the context, the entire explanatory load of the accident – as well as the moral responsibility – is placed on the individual (human) component, which takes away all the necessity and opportunity to learn from the event. The systemic conditions that gave rise to the error are left in the system for the next person to run into. Besides being impractical, is it ethically right to put the next person in the same situation or system?

Automation

One common method to deal with 'human error' in socio-technological systems is by removing the human component from critical operational processes in the system. The 20th century has seen a tendency to replace humans by technology, from mechanisation to computerisation and automation – systems being controlled by autonomous technology rather than humans. Initially the introduction of automation had little to do with the performance variability that humans display – it was driven by the sheer possibility of technological progress and the desire to reduce costs through the reduction of human staffing requirements – but later the focus of innovation turned to technology as a manner to remove human error. Computers were to replace as many human tasks as possible in order to remove human operators from the control loop; humans were now tasked to simply monitor the automation.

Even if automation is more reliable than human operators, it does not remove the potential for 'human error' but merely relocates it. Even in highly automated systems humans continue to be a central element, as they need to install, monitor, maintain, and eventually improve the system. Bainbridge (1983) suggested that the increased reliance on technology, ironically, also increases the importance of the human operator in various ways. First, the human operator's task has increasingly become that of monitoring the automation, even though human factors research has shown that the properties of human attention are generally ill-suited for (long-term) vigilance tasks. Second, automation may decrease operators' physical workload at the price of increasing their cognitive workload, especially during periods in which cognitive workload is already high. Third, as operators have become detached from their original work, when in the case of an emergency the automation fails, they may have lost most of their skills to perform the actual work but are now supposed to take over (from the automation) during highly critical and difficult moments. Replacing human work by technology does thus not simply reduce human error, but changes the nature of work and has some unanticipated consequences.

Not only does automation change the nature of work, it also does so in a manner that can be seen as practically and ethically problematic. In addition to the unforeseen effects that come with the introduction of new technology, one important ethical issue that deserves some elaboration here is the question of who gains and who loses by the ubiquitous introduction of automation. As many human tasks and roles have been replaced by technology, a great number of human-human interactions have been replaced by human-machine interactions. This severely limits the options for people to have meaningful social interactions at work. Similarly, the pride and joy people took in their traditional tasks and roles has been replaced by the automation doing their work for them; leaving the human operator to monitor the automation. Should we automate as many tasks as possible, as well as require operators to maximise their use of automated modes (think about pilots managing the autopilot rather than flying the airplane), just because technological progress allows us to do so or because it provides for more predictability and productivity for management? Is this serving the well-being of the individual operator? Is the broad and ubiquitous introduction of automation in the best interest of society?

The intention to remove the potential for 'human error' is of course supportive of a basic aim in human factors, and entirely in line with Fitts and Jones' approach for the human-centred design of interfaces. However, the intent to simply replace the human component by automation does not constitute a systems approach. It builds on the reductionist assumption that assumes the system works fine if it were not for the unreliable human components. This rationale, however, overlooks the central role that humans play in enabling the system to exist and function in the first place. It overlooks the fact that humans, more specifically human improvisation and performance variability, are the only resource that can balance goal conflicts and adapt to the various unexpected states that the larger system throws at them. A systems approach would see the system and automation as constituent of the human – as humans we use tools to augment, supplant, and aid ourselves – as well as the human being constituent of the system. As such, a systems approach concerning automation might consider the (possibly less efficient) option of having the automation monitor the human. In the traditional approach of humans monitoring the automation, human factors has changed the nature of work in the system, but not necessarily for the better.

Procedures

To deal with the problem of incidents and accidents in our systems, reductionism dictates that if all parts function as they are supposed to, so will the whole. As such, probably most indicative of human factors' reductionist approach is its solution to deal with the human performance variability in a similar manner as it would deal with technical components: make the components more reliable by tightly controlling their behaviour. Proceduralisation provides operators in complex systems with precise instructions (often presented in the form of checklists) on how to resolve both normal and emergency situations and so increases the reliability and predictability of their performance. This kind of performance standardisation through proceduralisation is not a new idea, nor unique to human factors. Breaking down complex work into smaller – but more manageable and efficient – parts can be traced all the way back to the work of Frederick Taylor in the early 1900s:

> "*Perhaps the most prominent single element in modern scientific management is the task idea. The work of every workman is fully planned out by the management … , describing in detail the task which he is to accomplish, as well as the means to be used in doing the work. … This task specifies not only what is to be done but how it is to be done and the exact time allowed for doing* it." (Taylor, 1967, p. 39)

As a management philosophy – with its principles applied to a broad range of industries and still popular today (e.g. Lean-production) – Taylorism assumes that by breaking down tasks and meticulously prescribing work, efficiency and reliability can be implemented from the top down. It provides the foundation for the belief that procedures (often specified by supervisors and middle-management) are the best way to specify tasks and roles (for operators). Most safety-critical domains rely on procedures and compliance to increase the predictability and reliability of people's performance. "*Over time proceduralisation has become more than an answer of how to increase safety in modern socio-technical systems, it may have become* the *answer*" (Bergström et al., 2009, p. 76, emphasis in original).

As a human factors safety solution, the reliance on procedures and compliance has its practical limits. Every time an unwanted event occurs, a new set of increasingly detailed procedures is introduced to respond specifically to just the latest incident. For example, after the friendly fire on two Black Hawk helicopters over Northern Iraq, "*higher headquarters in Europe dispatched a sweeping set of rules in documents several inches thick to 'absolutely guarantee' that whatever caused this tragedy would never happen again*" (Snook, 2000, p. 201). Over the years, this has resulted in a suffocating level of procedural over-specification, making it impossible for people in almost any system to comply with all the rules. This can clearly be seen in 'work-to-rule' strikes in which people take every single procedure in the rule-book in its full and literal meaning, often resulting to gridlock production. We need to ask ourselves, if it is ethically right that operators routinely need to work around or loosely interpret many official procedures – a phenomenon discussed in human factors as the gap between 'work-as-imagined' and 'work-as-done' – to get their work done? Additionally, as the current level of proceduralisation has made the system increasingly opaque, we need to ask whether safety is served by making it very difficult for the people inside the system to understand what the important rules of the system are?

Of course, everybody in society profits from an efficient production machinery, aided by automation and highly efficient ways to manage work. It has brought the (Western) world great prosperity. However, this manner of organising work seems to lose track of human factors' aim of increasing human well-being. Just as with replacing human work by automation, breaking complex tasks down into smaller more efficient sub-tasks, to be performed by different people, makes work less meaningful. Instead of constructing an entire bicycle, factory employees are now assigned the task of fitting just one specific part to the bike-frame, and they are supposed to do so thousands of times a day. Task- and labour-specialisation, as well as an increased reliance on prescribed behaviours, do not only reduce the meaningfulness of people's work, but may also reinforce people's alienation from work and the loss of professional identities.

Moreover, tightly prescribing work may contribute to making our work-force dumber and less likely to come up with (creative) solutions to problems. This is particularly problematic as procedures are necessarily incomplete specifications for action – they are always simplifications of the real world and can never foresee all scenarios in a complex system. The abilities to recognise, absorb and adapt to disruptions that fall outside the system's design base have been identified as the key aspects of making individuals, and consequently the organisations that they make up, resilient (Hollnagel, Woods, & Leveson, 2006). However, can we expect this kind of resilience from operators whom we have instructed not to think for themselves, to be 'docile bodies'? As such, our reliance on procedures and automation comes with the necessary sacrifices, both practically and ethically.

Conclusion

The instances above – practices in regards to taxonomies, automation, and procedures – only represent a small number of the issues that human factors is concerned with. We could also have argued for the need to re-think our insistence on the (self-) reporting of incidents: is it right to ask people to report on their own or other people's behaviours, especially if this will not be judged in light of the complexity of the context in which work takes place? We could have raised our ethical concerns with the way notions such as 'resilience engineering' are sometimes used: basically asking people to accept more risk by relying on their individual abilities to cope with danger and surprise, when no systemic changes are 'engineered'. Other examples abound. The larger point is, however, that despite the work of pioneers as Fitts and Jones, who claimed that the features of people's equipment and environment shape human performance, we largely remain wedded to a reductionist focus on the human component of our systems.

Of course there is nothing wrong with reductionism per se. It becomes problematic – both practically and ethically – when it is taken as the only approach. As a human factors community, we need to ask whether the explanations, labels, classifications, and interventions we propose are serving our purposes. Given the complexity of socio-technological systems, adverse outcomes may no longer be satisfactorily explained by reference to individual components and cause-effect relations. A systems perspective might be better equipped to help us understand how the system shaped the behaviour of those in it and will be able to make lasting changes to prevent future incidents. As the complexity of our socio-technological systems has reached a point where many of these systems can no longer be modelled or controlled, it may thus be increasingly irrational to keep relying on the reductionist approach, characteristic of our discipline. As we have tried to illustrate above, continuing to do so comes with a price.

If we cling to our commitment to find broken parts, as a modern way of scape-goating, we are asking for ethical trouble with our human factors explanations and practices. As long as the self-contained individual remains the unit of analysis for human factors, we are at not helping other discourses, such as the juridical, to move away from accounts of individuals as the fundamental elements of thought and action. We are not doing anything to point them to explanations and vocabularies of complexity and systems – to models of accountability in which systems as a whole break down, not their substituent parts. Constructs, such as SA, which we once introduced to help operators better understand and deal with complex dynamic environments, are now turned against them as they provide normative standards for behaviour. Even if SA is a useful human factors construct, the manner in which other discourses take and use it to ask (and answer) questions about error, blame, and criminality is an ethical problem that we as a human factors community need to take responsibility for. Perhaps, in a small but not unimportant way, we are complicit in Karl Lilgert's sentence, and in him sitting in jail today.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Bergström, J., Dahlström, N., van Winsen, R., Lützhöft, M., Nyce, J., & Dekker, S. (2009). Rule-and role-retreat: An empirical study of procedures and resilience. *Journal of Maritime Research, 6*(3), 41-58.

Dekker, S.W.A. (2011). *Drift into failure: From hunting broken components to understanding complex systems*. Farnham, England; Burlington, VT: Ashgate.

Fitts, P.M., & Jones, R.E. (1947). *Analysis of factors contributing to 460 'pilot error' experiences in operating aircraft controls.* Wright-Patterson Air Force Base, Dayton, Ohio: Aero Medical Laboratory, Air Material Command.

Galison, P. (2000). An accident of history. In P. Galison & A. Roland (Eds.), *Atmospheric flight in the twentieth century*. Dordrecht; Boston: Kluwer.

Hollnagel, E. (2006). Accident analysis and "human error". In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (pp. 1889-1892). Boca Raton, FL: Taylor & Francis.

Hollnagel, E., Woods, D.D., Leveson, N. (2006). *Resilience engineering: Concepts and precepts*. Burlington, VT; Aldershot, England: Ashgate.

International Ergonomics Association. (2014). Definition and domains of ergonomics. Retrieved April 19, 2014, 2014, from http://www.iea.cc/whats/index.html

Keller, J. (2013). Karl lilgert on trial for bc ferry queen of the north sinking, *The Canadian Press*. Retrieved from http://www.huffingtonpost.ca/2013/01/17/karl-lilgert-bc-ferry-queen-of-the-north-sinking_n_2493214.html

Meister, D. (1999). *The history of human factors and ergonomics*. Mahwah, N.J: Lawrence Erlbaum Associates.

Reason, J.T. (1990). *Human error*. Cambridge [England]; New York: Cambridge University Press.

Snook, S.A. (2000). *Friendly fire: The accidental shootdown of U.S. black hawks over northern Iraq*. Princeton, N.J.: Princeton University Press.

Supreme Court of British Columbia. (2013). *Regina v. Karl-heinz arthur lilgert*. (2013 BCSC 1329). Vancouver.

Taylor, F.W. (1967). *The principles of scientific management*. New York: Norton.

Transportation Safety Board of Canada. (2008). *Striking and subsequent sinking, passenger and vehicle ferry queen of the north, gil island, wright sound, british columbia, 22 march 2006*. (M06W0052). TSB.

Appendix B.

SA Anno 1995: A commitment to the 17th century

This chapter includes a co-authored paper. The copyright of this article has been transferred to SAGE, however, authors have the right to include their article in a thesis or dissertation, provided that this is not for commercial purposes.

The bibliographic details of the co-authored paper, including all authors, are:

Winsen, R. van, & Dekker, S. W. A. (2015). SA Anno 1995: A commitment to the 17th century. Accepted by the *Journal of Cognitive Engineering and Decision Making* (August 22, 2014).

As a the first-author of this paper I have had a leading role in pursuing this argument; I have reviewed the literature on SA; I have written the 'rebuttal' to Endsley's article that serves as the basis for the special issue in which this paper is published; I have implemented the feedback from the co-author; and I have implemented the feedback from the editor.

(Signed) ______________________ (Date) 22-02-2015
Roel van Winsen

(Countersigned) ______________________ (Date) ____________
Sidney W. A. Dekker (Principal supervisor and co-author)

SA Anno 1995: A commitment to the 17^{th} century

Abstract

In providing reflections on 25 years of SA research, particularly the ever-popular 1995 model of SA, our response is twofold. First, we ask whether the model's grasp has exceeded its epistemological reach. By overlooking important insights from the second cognitive revolution as well as from other late twentieth-century developments in (social) science, it might well do that. In fact, SA, in its 1995 model, is strongly committed to a 17^{th} century ontology that separates mind from matter and sees awareness as a correspondence or mirror of the world outside. This seems to strongly reverberate today in a somewhat dogmatic stance of the 1995 model about the role that the world and cognitive artefacts play in constituting cognitive processes. Second, we suggest that, after 25 years of SA, we might need to reflect on what SA as a scientific human factors object has brought us, or the operators we once set out to support. This is not a trivial or academic question. We know of one operator who is in jail today because the prosecution was able to successfully argue for the dereliction of his duty to maintain SA. Without the HF community supplying this object, he might still be in jail, but surely not under this charge.

Introduction

In reading the contributions to this issue, it is heartening to see the deliberations on what we, as a human factors community, have to offer operators, designers and others working with complex, dynamic systems. And indeed, we should all welcome opportunities to jointly assess whether our grasp exceeds our epistemological reach. That is—do we put more faith in some of our own ideas and models than is warranted? Epistemology asks how we know what we know. If we claim to know a lot about our subjects' awareness on the basis of one model of SA (and its assorted measurement techniques), those claims should obviously be subject to some examination. As they are here. Faith in our models sometimes derives from their popular adoption and large bodies of empirical evidence generated from that. But this has little to do with inherent scientific quality. That the widespread application of a model is able to deliver facts about the existence of its own constructs is not at all persuasive—at least not since Kuhn (1962). After all, most human factors constructs can claim to be supported by an ample empirical record—whether SA, mental workload, or automation surprises. The empirical record becomes available, after all, because the models themselves generate the demand for the data and the criteria by which their evidence is defined. Thus, ever-greater accumulation of facts does not equate with scientific quality, and certainly not with innovation or scientific evolution.

The 1995 model of SA represents a 17th-century ontology. Ontology is the branch of philosophy that inquires after the nature of being in the world—and for SA, as for 17th century thinkers like Rene Descartes and (a bit later) Isaac Newton, such being in the world means having an accurate representation of the world in the mind. A separation of mind and matter (or head and world) is fundamental to this view, and it has dogged cognitive science well into the 20th and even 21st century (Norman, 1988). Information processing (roughly popular from the 1950's to the 1970's in cognitive psychology), for example, has faithfully replicated 17th-century ontology. SA, too, is that correspondence with the environment; it is that mirror, that mimic, that mental simile of the world outside. Protestations to the contrary—"SA … goes beyond traditional information-processing approaches in attempting to explain human behavior in operating complex systems" (Endsley, this issue)—do not make it less so. SA is grafted onto the exact language, concepts, and flow charts of traditional information processing models. And indeed, with such 17th century dualism as its starting point, notions of "complete" or "accurate" SA become sensible and intelligible. For SA, "there is a "ground truth" against which its accuracy can be assessed (e.g., the objective state of the world or the objective unfolding of events that are predicted)" (Parasuraman, Sheridan, & Wickens, 2008, p. 144). It is an ontological commitment with consequences, however. It makes sacrifices, both epistemological and ethical, that deserve some reflection.

Dynamics and the environment

Contesting the critique that the model is not dynamic, the argument is that the "[m]odel shows a dynamic feedback loop for gathering information and acting on the environment" (Endsley, this issue). This constitutes a 1950's cybernetics approach to operationalising dynamism. The basics of interaction are there, but there is no escape from the static, linear, reductionist arrows and boxes. Since a decade or two, the human factors community has largely committed to the second cognitive revolution (the first was the move away from behaviourism and into information processing in the 1950's), which sees the environment as an active determinant of and participant in cognitive processes. SA, as John Flach (1995) would say, is largely about the A, and much less about the S. If "not a 'strictly in the head' model" (Endsley, this issue), it surely is a 'largely in the head' model (the environment, or human interaction with it, is depicted as a single arrow flowing from 'State Of The Environment' towards the SA 'box', see Endsley's figure 1, this issue). The major preoccupation—theoretically, methodologically—is a subject's *awareness*. This focus sacrifices our study of the environment, the *situation*, as both constitutive of and constituted by that awareness.

The commitment to awareness 'in the head' is so strong that Hutchins' (1995) work, as that of others in Joint Cognitive Systems (Hollnagel & Woods, 2005; Woods & Hollnagel, 2006), is deemed not heretical but simply absurd. Data can be all over a situation, but awareness is exclusively in the head: "I do not call such data, residing in a report or a display or an electronic system, 'situation awareness.' Inanimate objects do not have 'awareness' of the situation or of anything else" (Endsley, this issue). Such privileging of human consciousness is not without philosophical or religious historical precedent. But what it seems to forget is the fundamental phylogenetic differentiation that occurred through tool use—we are human not just because we are aware, but because we use tools to augment, supplant and aid our cognition, our awareness. Joint cognitive systems thinking tries to take that into better account than 'in the head' models of cognition ever could. Human cognition both constitutes and is constituted by operational, design- and organizational features of work, which suggests that human-machine interaction (the system) should be the unit of analysis (not the inside mapping of the environment in an operator's head). Salmon et al (2008), suggest how such a JCS approach to SA differs from the 1995 SA model:

> The main difference between individual and team models of SA and DSA [Distributed SA] approaches relate to the treatment of SA as a cognitive construct or as a systems construct. Individual and team models suggest that SA exists in the mind of individuals, whereas DSA approaches view SA as an emergent property or a product of the system itself. (Salmon et al., 2008, p. 313)

In reaction to such a systems approach, the conclusion is that "the implication of describing SA as 'an emergent system property' that is distributed across the system is that it essentially becomes sufficient as long as the needed SA is distributed somewhere in the system" (Endsley, this issue). This of course misconstrues a joint cognitive system, but also the notion of emergence. By taking the (joint cognitive) system as the system that is supposed to have a particular level of SA, instead of its individual components, we can regard SA as something that 'emerges' during and from cognitive work. Alas, the defence of 17th century ontology goes on: "Salmon et al. confuse sources of SA (displays, computers, and other artifacts) with SA itself, stating that the SA is in the artefact. This would only hold true if the artifact were itself a cognizant and independent decision maker" (Endsley, this volume).

From this, however, only more confusion follows, and beyond it not much else: "Information that exists in the environment, but which the decision maker is not aware of is by definition not situation awareness" (Endsley, this volume). The definition of SA becomes increasingly circular. What about information that exists in the environment—is it even information at all unless it is perceived or conceived of as information by some kind of 'cognizant and independent decision maker'? In that case, we do not need SA, or its arrow feeding in from the environment. It also no longer makes sense to talk about accuracy or completeness of SA (as its adherents like to do). The point, not just epistemological, but also practical, is probably to try to understand how people perceive the environment in ways that support action and goal achievement (Weick, 1995). On that, it is unfortunate that broader conceptualizations of 'information' and 'awareness' are dismissed out of hand (even though the importance of goal-driven or top-down processing is acknowledged). It slims allowable cognition down to what the 1995 model dictates. "While operators can and often do create heuristics and reminders in their work environment to assist them in keeping up with information and their task status, in my view these work methods and cues do not constitute SA" (Endsley, this issue). Retaining a narrow definition of SA just because the model says so—and in the face of much evidence for the critical role of heuristics and recognition in dynamic work (Klein, 1998), or see the notion of "level IV evidence" in medical practice (Stuebe, 2011)—is a bit dogmatic, and probably unhelpful for the evolution of our science.

The ethics of the effects of SA

The scientific object of SA has really taken off over the last 25 years. Situation awareness—and the loss of it—is everywhere: there are books, articles, journals, conferences, training courses, operating manuals, design guidelines, and accidents investigation reports addressing it. The human factors community once embraced a specific and defined object of SA to help with the design of human machine interfaces and to explain behaviour in complex dynamic environments. But is the object of SA also used in this way today? And who is responsible for the effects if it isn't? As a human factors community, it is imperative that we engage in ethical discussions, asking questions about the consequences of the explanations, classifications, and practices we propose.

The wide use of SA has not only reinforced its status—the more it was used, the greater the consensus authority on using it—but also driven a gradual move away from its original purpose. It has offered new normative standards for behaviour: pilots now describe themselves as good pilots when they are situationally aware, when their decisions are informed by 'good SA'. Soon after its conception, SA started to appear in accident investigation reports where it was given great causal power to explain accidents (see, for example, Aviation Safety Council, 2002). It has even showed up in courts. Operators get convicted for not living up to their deontological duty to (always) maintain situation awareness (Supreme Court of British Columbia, 2013). Even if SA is a 'useful' construct in human factors, the manner in which other discourses—most prominently the juridical—take and use it to ask (and answer) questions about error and culpability is an ethical problem that we as a human factors community need to acknowledge some responsibility for.

As long as the self-contained individual remains the unit of analysis for understanding situation awareness (or other aspects of cognition), we are at least not helping other communities (like the juridical) move away from 17th-century characterizations about individuals and individual minds as the fundamental elements of thought and action. We are not doing anything to point these other communities to languages of systems, to new models of accountability and responsibility that are much better adapted to the flatter, networked societies that are so prominent and rapidly expanding in the 21st century. In fact, by supplying objects like SA, which re-affirm a 17th-century ontology, we might even be abetting them. As long as it is human factors orthodoxy to reduce safety problems to their component parts (either human or technical), and to separate what was in a world (which adherents assert is "objectively apprehensible") from what was in the mind, we are asking for ethical trouble with our models and concepts.

In reflecting on 25 years of SA, we hope that human factors, and with it SA research, is ready to move into the twenty-first century by taking the lessons from the twentieth at heart. The time has come to move away from our commitment to individualism, reductionism, and dualism, and to move instead towards studying work in joint cognitive systems.

References

Aviation Safety Council. (2002). Crashed on a partially closed runway during takeoff, singapore airlines flight 006, boeing 747-400, 9v-spk, cks airport, taoyuan, taiwan, october 31, 2000. (ASC-AAR-02-04-001). Taipei, Taiwan: ACS.

Flach, J.M. (1995). Maintaining situation awareness when stalking cognition in the wild. Paper presented at the International Conference on Experimental Analysis and Measurement of Situation Awareness.

Hollnagel, E., & Woods, D.D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. Boca Raton, FL: Taylor & Francis.

Hutchins, E. (1995). Cognition in the wild: MIT Press.

Klein, G.A. (1998). Sources of power: How people make decisions. Cambridge, MA: MIT Press.

Kuhn, T.S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Norman, D.A. (1988). The psychology of everyday things. New York: Basic Books.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. Journal of Cognitive Engineering and Decision Making, 2(2), 140-160.

Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., Jenkins, D.P., McMaster, R., & Young, M.S. (2008). What really is going on? Review of situation awareness models for individuals and teams. Theoretical Issues in Ergonomics Science, 9(4), 297-323.

Stuebe, A.M. (2011). Level iv evidence — adverse anecdote and clinical practice. The New England journal of medicine, 365(1), 8-9.

Supreme Court of British Columbia. (2013). Regina v. Karl-heinz arthur lilgert. (2013 BCSC 1329). Vancouver.

Weick, K.E. (1995). Sensemaking in organizations. Thousand Oaks: Sage Publications.

Woods, D.D., & Hollnagel, E. (2006). Joint cognitive systems: Patterns in cognitive systems engineering. Boca Raton: CRC/Taylor & Francis.