The Quest for Edge Awareness

Lessons not yet learned

PhD Thesis

on practical and situated usefulness of advanced technological systems
among inescapable uncertainties and competing interests
in a world of dynamic changes

by Patrik Stensson

Dissertation presented at Uppsala University to be publicly examined in Sal IX, Universitetshuset, Biskopsgatan 3, Uppsala, Monday, 8 December 2014 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor emeritus Erik Hollnagel (Syddansk Universitet).

Abstract
Stensson, P. 2014. The Quest for Edge Awareness, Lessons not yet learned. PhD Thesis on practical and situated usefulness of advanced technological systems among inescapable uncertainties and competing interests in a world of dynamic changes. 297 pp. Uppsala: Institutionen för informatik och media. ISBN 978-91-506-2425-0.

This thesis problematizes the concept of usefulness, in part by taking questions to the extreme. The starting point is the contemporary view of usefulness, a view that remains within a traditional paradigm of technical rationality in which important aspects are disregarded or not perceived because they are not part of the equation. For scrutiny of technological usefulness that is a socially situated phenomenon regarding physical systems, neither interpretivist nor positivist research approaches are sufficient. Both views are required. Critical Realism supports such duality, facilitating the combination of elements from different paradigms, and provides methodological guidelines for doing this. The critical realist approach makes it possible to transcend the boundaries of technical rationality and contribute an alternative definition of usefulness that takes into account also the situated, the contextual, and the unpredictable. The aim is that this definition will contribute to a transformation of society.

Concepts related to usefulness, such as predictability, controllability, effectiveness, and safety, are revisited, redefined, or complemented. Underlying aspects and mechanisms are explored and tensions identified, resulting in a theoretical contribution with models and frameworks explaining what is argued to be the true nature of usefulness. Potentiality is suggested as a complementary concept to effectiveness, similar to how resilience complements safety. Situated usefulness is then defined using these four concepts. The phenomenon known as situation awareness is scrutinized as well, and complemented by system awareness and the thesis title concept, edge awareness.

Four cases, two airline crashes and two nuclear power plant events, and three future scenarios, constitute the empirical contribution. The analysis shows that the contributed frameworks and redefinition of usefulness facilitate different or extended explanations of all four events, and that future cases lack considerations of situated usefulness. Research implications center on the human role and our responsibilities in relation to the technology that we use, and on the meaning of concepts defining this role. We are situated human beings. Our role is to be involved and responsible, a role requiring awareness and controllability. The escalating ubiquity and the character of computerized technological systems make therefore the quest for edge awareness more important than ever.

Keywords: critical realism, human factors, situation awareness, resilience engineering, human-machine interaction, automation, autonomy, situated usefulness, systems thinking, sociotechnical systems, edge awareness

Patrik Stensson, Department of Informatics and Media, Kyrkogårdsg. 10, Uppsala University, SE-751 20 Uppsala, Sweden.

ISBN 978-91-506-2425-0
urn:nbn:se:uu:diva-234369 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-234369)

Till Eva, Julia, Albin och Alma
Det är för er skull jag försöker förbättra världen

Contents

1 Introduction 1
1.1 Outlining the research topic 2
1.2 Research objectives 7
1.2.1 General research setting 8
1.2.2 Research questions 10
1.2.3 Structure of thesis 10
1.3 Setting the frame of reference 14
1.3.1 Edges are everywhere 14
1.3.2 Usefulness, of what? 16
1.4 Advanced technology, for better or worse? 17
1.4.1 The categorical limit for technology 19
1.4.2 Automated safety and skidding cars 20
1.4.3 Safety on a plateau, from two perspectives 22
1.4.4 The brittle safety with modern airliners 24
1.4.5 The analogy between safety and usefulness 26

2 Research approach 29
2.1 Philosophy 29
2.1.1 Critical realism 30
2.1.2 Openness, explanations, and the epistemic fallacy 34
2.1.3 Stratification and emergence 36
2.1.4 Knowledge and practical wisdom 39
2.1.5 Two perspectives, detached and involved 43
2.2 Methodology 45
2.2.1 Critical realist methodology 45
2.2.2 On problematization and theorizing 46
2.2.3 The nature of theory 48
2.2.4 Prescience 50
2.2.5 Making experiences and preconceptions explicit 51
2.2.6 Relevance to practice 53
2.3 Research method 54
2.3.1 Phase one – theory generation 55
2.3.2 Phase two – case studies 58
2.3.3 Phase three – discussions 60

3 Experiences and preconceptions 61
3.1 Training for edge awareness 63
3.2 Pilot engineer 66
3.3 Human factors in aviation, and in car driving 69
3.4 Desired end-states in the fog of war 75
3.5 Military-technology vs. military technology 82

– Phase one – 83

4 Outline of phase one results 85
4.1 Models and frameworks, R1 85
4.2 Explaining usefulness, R2 88
4.3 The vicious circle culture, R3 89

5 Exploring usefulness 93
5.1 Systems thinking 94
5.2 Sociotechnical worldview 98
5.3 Character of technological systems 104
5.4 Character of computerized systems 107
5.5 Summary of worldview models 109

6 Desired effects 111
6.1 Critical systems thinking and discourse 113
6.2 Character of utility 114
6.3 From conditions to effects, in theory 116
6.4 Summary of value models 120

7 Uncertainty and unpredictability 121
7.1 Safety science 122
7.2 Character of uncertainty 125
7.3 Effectiveness and safety 128
7.4 Potentiality and resilience 130
7.5 Summary of difficulty characteristics 132

8 Models and scenarios, and simulators 133
8.1 The human contribution 134
8.2 The character of controllability 136
8.3 From conditions to effects, in practice 137
8.4 Summary of practice models 139

9 Situation awareness 141
9.1 The human component 142
9.2 Situation awareness and system awareness 146
9.3 Edge awareness 149
9.4 Summary of psychological concepts 155

10 Explaining usefulness 157
10.1 Conceptual framework for usefulness 157
10.2 The character of usefulness 159
10.3 The case of design and concept meanings 167

– Phase two – 169

11 Case studies, exploring RQ2 and RQ3 171
11.1 Case study methodology 171

12 Cases from the past and the present 175
12.1 SAS, SK751, December 27th, 1991 175
12.1.1 Accident investigation findings 179
12.1.2 Interpretations according to present frameworks 182
12.2 Forsmark, July 25th, 2006 186
12.2.1 Report findings 189
12.2.2 Interpretations according to present frameworks 191
12.3 Air France AF447, June 1st, 2009 195
12.3.1 Accident investigation findings 201
12.3.2 Interpretations according to present frameworks 204
12.4 Fukushima, March 11th, 2011 208
12.4.1 Accident investigation findings 212
12.4.2 Interpretations according to present frameworks 215

13 Future technologies 221
13.1 Augmented reality 221
13.1.1 Report findings 222
13.1.2 Interpretations according to present frameworks 224
13.2 Nano air vehicles 227
13.2.1 Report Findings 227
13.2.2 Interpretations according to present frameworks 228
13.3 The MODAS Project 229
13.3.1 Current investigations and continuation of the project 230
13.3.2 Interpretations according to present frameworks 231

14 Phase two conclusions 233
14.1 Addressing RQ2, the case of use 233
14.2 Addressing RQ3, the case of design 237
14.3 General conclusions 239

– Phase three – 241

15 Research implications 243
15.1 The meaning of concepts 243
15.1.1 Intelligence, autonomy, and other human concepts 244
15.1.2 The current use of human concepts for technology 246
15.1.3 The vicious circle of drifting concept meanings 248
15.1.4 Six sources of concept confusion 249
15.1.5 Consequences of concept confusion 253
15.2 Human authority, a categorical imperative 255
15.3 Philosophical reflections 259
15.3.1 Scientific abolition of the generative present 260
15.3.2 Impact on the human role 262
15.3.3 Impact of computerization 263

16 Discussion 265
16.1 The logic of discovery 265
16.2 The core insight – true usefulness is situated 268
16.3 Situated usefulness explained 270
16.4 Validity of contributed models and frameworks 272

17 Conclusions 277
17.1 Future research directions 280

18 Summary in Swedish 281

Figures

Figure 1.1: Thesis structure....11
Figure 1.2: SAS SK751 wreckage at Gottröra, a crash caused by ice breaking off from the wings damaging the engines, or, is inappropriate human interaction with automated systems a more plausible explanation? (Photo: SHK)....13
Figure 2.1: Aristotelian-inspired categorization of human virtues....41
Figure 2.2: Conceptual design of research – phase one....56
Figure 2.3: The relation between theoretical areas and described experiences....57
Figure 2.4: Conceptual design of research – phase two....59
Figure 3.1: An example of a simulator for both visual and motion cues. (Source: wikimedia, public domain)....79
Figure 5.1: sociotechnical system outline (R1.1.1), the worldview....99
Figure 5.2: The recursive nature of technological properties (R1.1.2)....105
Figure 5.3: System integration before and present....108
Figure 6.1: From conditions to effects (R1.1.5), in theory....117
Figure 8.1: From Conditions to Effects, in practice (R1.2.3), i.e., walking the map....137
Figure 9.1: Two modes of human thinking, adopted from (Kahneman 2003, p. 698)....145
Figure 9.2: Typical friction and lift curves, examples of continuous performance relations with (intuitively) observable extremes....151
Figure 9.3: Situation Awareness, System Awareness, and Edge Awareness (R1.2.4)....154
Figure 10.1: Conceptual framework for usefulness (R2.1)....157
Figure 10.2: The character of usefulness (R2.2), with two tensions as its dimensions....160
Figure 11.1: Relative layout of cases studied....172
Figure 15.1: The vicious circle of reduced concept richness....248
Figure 15.2: Philosophical assumptions leading to the lost situated perspective....261
Figure 16.1: Character of usefulness, the logic of discovery....268
Figure 16.2: Contributed models and frameworks, explanations, reasons for plausibility, and foundations for reasoning....273

Figures 5.2, 6.1 and 15.1, have been published in *Ergonomics,* 57 (3), they are included with permission from Taylor & Francis.

Figure 9.3 has been published in proceedings of the AHFE2014 conference.

Tables

Table 2.1: The domains of critical realism.....33
Table 14.1: Summary of the four cases, part 1: Situation Awareness.....233
Table 14.2: Summary of the four cases, part 2: System Awareness.....234
Table 14.3: Summary of the four cases, part 3: Edge Awareness.....234
Table 14.4: Summary of the four cases, part 4: Conclusions.....235

Separate publications

Stensson, P., 2010. Thoughts about the Consequences of Inappropriate Application of Systems Engineering. In: Proceedings of the 7th European Systems Engineering Conference. Presented at the EuSEC 2010, Stockholm: INCOSE.

Stensson, P. and Jansson, A., 2014a. Autonomous technology – sources of confusion: a model for explanation and prediction of conceptual shifts. Ergonomics, 57 (3), 455–470.

Stensson, P. and Jansson, A., 2014b. Edge Awareness - A Dynamic Safety Perspective on Four Accidents/Incidents. In: B. Amaba and B. Dalgetty, eds. Advances in Human Factors, Software and Systems Engineering. Krakow: AHFE Conference, 168–179.

Jansson, A., Stensson, P., Bodin, I., Axelsson, A., and Tschirner, S., 2014. Authority and Level of Automation: Lessons to be learned in design of in-vehicle assistance systems. In: M. Kurosu, ed. Human-Computer Interaction. Applications and Services. Springer International Publishing, 413–424.

Central concepts

Awareness, Situation/System/Edge: Psychological phenomena. Situation awareness is in this thesis specified to denote awareness about the situation in which the individual having the awareness and the system of concern is situated. Situation awareness is about what is going on in the environment. System awareness is about what is going on in the technological system of concern. Edge awareness is then defined as the situated synthesis of situation and system awareness, an awareness about dynamic interaction effects between the system of concern and its environment. Typical interaction characteristics that comprise edges are described by the friction curve (e.g., for cars) and the aerodynamic lift curve (for aircraft). For human beings to know (intuitively) about system effects, to understand (emotionally) implications of system control, and to judge (insightfully) the contextual values of system effects, awareness of such edges is argued to be essential.

Calculable/Calculative: Possible to describe by mathematics. A calculative approach is to consider things describable, derivable, and predictable by mathematics, which may be appropriate for certain kind of matters, such as matters from the world of Newtonian physics. However, for other matters, not appropriately represented by mathematical states, the application of a calculative approach is in this thesis argued to cause problems.

Complexity/Complicated: Complicated indicates a difficult problem to solve in a concrete (calculable) world. Complexity indicates a situation where it is difficult to find concrete problems. Complexity comes from multiple, non-linear, multivalent, circularly dependent, ambiguous, vague, etc., relations and dependencies. In particular, complexity may come from subjectivity and clashing desires based on incomparable aspects. Complexity is thereby inherent in systems comprising autonomous human beings.

Context: In some sense a more abstract version of situation, denoting a social and cultural situation, a position within a discourse, forming the conditions for meanings. As such, it is often quite difficult to delimit 'the system' for which a context is relevant, or the other way around, to delimit the context relevant for 'a system'.

Deterministic: Ties in with calculable/calculative in the context of systems. A formally describable monovalent system is deterministic, meaning that a certain state determines a following state, or a certain input determines a specific and unambiguous output.

Edge: An edge implies a distinct shift of some kind. Physical edges are the most straightforward ones and the cliff edge is a very illustrative metaphor. However, the kind of edges mostly referred to in this thesis is abstract edges in relations between aspects, although they are often edges with tangible physical effects if they are crossed. Performance edges of different kinds are the primary focus. These edges often occur between physical properties of the system of concern and its environment (e.g., the car and the road) and they are often dynamic in their nature. That is, tempo, situation (over some time), context, etc. tend to affect the character of the edge.

Effectiveness/Efficacy/Efficiency: To what extent something is effective. Efficacy most often denotes effective *for what*, while efficiency denotes *how much* effective something is, for example in relation to required resources. Effectiveness is similar to efficiency, but relating to less concrete resources, more in terms of *how* effective something is for a certain purpose. However, all these concepts tend to be used in place of each other. In this thesis effectiveness is the most common term, predominantly used with the intention of covering effective for what purpose, i.e., a mixture of both efficacy and effectiveness as these concepts are outlined here.

Experience/Experiences: Sometimes the word experiences is used to denote scientific observations when there is a need to stress their subjective nature, occasionally also throughout this thesis. However, experience may also denote an individual's (practical) knowledge, which in some sense is an aggregation and blending of life-experiences, but inescapably also the result of interpretation and reflection thus not merely a cumulative collection of concrete input. Experiences may therefore also mean a number of different kinds of experience. The intended meaning is hopefully possible to understand from context.

Framework: A more loosely coupled or abstract kind of model, or a set of tools for depicting or describing a model or phenomenon, e.g., a conceptual framework. In this thesis, the word framework is mostly used to designate descriptions about how models or concepts relate to each other in a general sense. Consequently, models may be considered a more specific kind of framework thus possible to be included when talking about frameworks in general terms. The word is here often used in that general sense.

Generative: Actively created, as opposed to passively determined. Within a fully deterministic system, effects are always the result of previous states. Generative effects are produced as the result of powers, such as deliberation and conscious intentions for human beings.

Model: A model aims to explain how certain things work, how a certain phenomenon comes to be. Models are always simpler than reality and limited in some sense, this is what makes them useful. The explanatory power of a model lies in its clarity and in the way it frames the explained phenomenon. However, there is always more to the world compared to the model and models risk thereby frame attention to modeled aspects and obscure other aspects. For physically oriented phenomena models are commonly descriptions of systems of one kind or another.

Paradox/Paradoxical: A paradox is something that seems impossible but in fact is true. If something is paradoxical, it happens despite that it should not happen. The latter notion may also be used slightly metaphorical, meaning that things are not what it appears to be.

Potentiality: A more generative (less calculative) form of effectiveness, especially regarding the selection of purposes. Efficacy, effectiveness, and efficiency, presuppose implicitly a fixed (unambiguously desirable) purpose according to which values are judged. Potentiality is about seizing opportunities, about selecting discovered, perhaps previously unknown, purposes that are judged as relevant and desired according to local values perhaps unique for that particular situation and context.

Resilience: A more generative (less calculative) form of safety. While safety is about being safe, resilience is more about becoming safe.

Safety: A complex concept discussed at length throughout the thesis, but it may be stressed here that the word safety is used to designate the absence of undesired effects in general. Thus, it is used in a sense that may incorporate several other related concepts such as security, assurance, protection, etc.

Situated: Present within a situation and context.

Situation: In some sense a more concrete form of context, forming the conditions for (concrete) matters, but still more, as in richer, than a state. In particular, a situation is here considered extending over some period, depending on the character of the system of concern.

State: A snapshot situation (or context), commonly represented by use of operationalized variables, e.g., measures.

Technical/Technology/Technological: Technical means of a technical nature, as in mechanical, detailed, requiring or following certain procedures, machine-like, etc. Technology refers to a specific kind of technological solution. Technological, on the other hand, is more generally referring to technology and aspects related to technological systems and technologies. A phrase like 'technical properties' risk being interpreted as 'properties requiring certain skills or techniques', as complicated properties, or as properties related to a specific domain. By talking about 'technological properties' instead, the intention is to put focus on properties and aspects related to the fact that there are technologies and technological systems involved. Hence, the phrase 'sociotechnical systems', which is used here as it stands, should with the above view and within the context of this thesis perhaps have been better phrased as 'sociotechnological systems'.

Tension: A tension is a rich concept indicating some kind of opposition between aspects. However, tensions are seldom absolute, meaning that they cannot be resolved. Issues implying tensions can coexist despite being contradictory. This is what creates the tension.

Theory/Theorizing: A theory is a more or less tentative explanation of things and theorizing is the process of producing this explanation. The strict notion of a theory is to denote an explanation usable for experimental predictions. However, the word theory is here often used in a general sense, as when talking about the theoretical contribution of this research.

Typology: Models and frameworks, as well as typologies, are all theories as in being some kind of explanation of things. It is the purpose of use and, perhaps, the process of creating these explanations that make them be either or. Typology is a kind of theory (model or framework) primarily intended to be used for analysis, not for explicating laws usable for predictions.

Uncertainty/Unpredictability/Equivocality: Describing to what extent something is unknown, from different perspectives. Uncertainty may come from lack of precision in measures, from lack of measures, or from lack of knowledge about missing measures, if viewed from a calculative perspective. Unpredictability may then be a consequence of uncertainty. For complex systems, systems with inherent tensions or contradictions, equivocality may therefore be a more appropriate word. However, uncertainty and unpredictability are most frequently used in this thesis.

1 Introduction

It does not matter if you like to keep a safe distance to every possible edge there is, if you get your kicks from 'living on the edge', or if you end up in a situation where you have to defeat an opponent by being more skilled in getting close to an edge. It is in any case of crucial importance to be thoroughly aware of the edge. Otherwise, you may unintentionally slip over an edge and fall helpless into whatever it means to be on the other side. Neither does it seem to matter whether you are a child or an adult, nor whether you are an organized professional or an unaffiliated individual. Today, regardless of your position, activities tend to include complex technological systems, and activities always comprise edges of some kind. Technological systems are ubiquitous and edges are everywhere. Consequences from issues with human-systems interaction and lack of *edge awareness* are therefore matters of concern for everybody.[1]

This thesis is about human beings in a high-tech society and about our role in relation to the technological systems that we develop and use, a relation much more complex than merely concerning people and artifacts interacting. How we look upon our role in relation to these systems shapes our view and forms our understanding of their usefulness, which in turn influences how we look upon our own role when using them, a circular dependency for better or worse. By applying critical scrutiny, this thesis aims to enhance our understanding of usefulness of modern technology.

For the sake of argument, the contemporary view of usefulness is described as governed by a still prevailing paradigm of industrial age technical rationality. It is a paradigm apparently favoring a model-based approach to meanings and values, an approach disregarding situated and contextual aspects. I will in this thesis argue that an exaggerated focus on predictability according to deterministic models and technical rationality make technological systems supposed to be safe and effective in fact become explicitly designed to provide *false safety* and *illusory effectiveness*, resulting in a poor kind of usefulness that from a human emancipation perspective also is counterproductive.

[1]As a general idea, in most cases where edge and edge awareness is mentioned, I recommend thinking about system performance edges such as the edge where your car starts skidding. The meaning of edge and edge awareness are, however, questions returned to several times throughout the thesis, the first occasion is already at the end of chapter 1.1.

The world is complex and to a great extent unpredictable, in particular regarding judgment of values, interpretation of meanings, and development of incentives. True technological usefulness requires therefore an ability to adjust system effects to suit local conditions and contextually relevant values. Moreover, some situations are more unpredictable than others are, military situations, for example. My professional background as a fighter pilot and Human Factors researcher provides thereby special input for critical scrutiny of technological usefulness because it has given me personal experience of using as well as of analyzing the use of advanced technological systems in situations where the edge between success and capital failure is tangible and where consequences are immediate. The practical and *situated usefulness* of technology in such extreme situations forms a model for an updated view of usefulness, a model or typology of usefulness characteristics that will be shown applicable to much more than the narrow field of military aviation. In short, the main purposes and objectives of this thesis are (for the more elaborate description, see ch. 1.2):

- To scrutinize the contemporary view of the concept of usefulness (of technological systems)
- To reflect on the impact that the prevailing view of technological usefulness has on human autonomy
- To contribute an alternative definition of usefulness that takes into account also:
  - the situated and the contextual
  - the unpredictable

1.1 Outlining the research topic

Do technological systems counteract the gaining of edge awareness? In this thesis, I will argue that some systems have this unfortunate effect, which, if that is the case, implies that technological 'structuralization' (Orlikowski 1992, drawing upon Giddens 1984) is strengthened by their designs. How can such counteraction happen, and what is the problem? Technological structuralization is strengthened, arguably, by a lack of edge awareness. When people as a consequence of systems design become unaware of how close to the edges they are, they begin relying on the technology to detect and recognize the edges, an adaptation that I believe should be much more of a concern for scientific discourse than what appears to be the case today. I will argue that as an intuitive and natural reaction of self-preservation, or comfort, people try to avoid the unpleasantness of falling over, or the extra effort required for climbing over, all kinds of edges, including technology-induced ones. Subconsciously we align our behavior to system-internal (software) models implemented as automations, templates, information

filters, predefined modes of operation, predetermined goals, and so forth. We adopt this kind of behavior in part because models and automations promise us safety, effectiveness, and efficiency. For persuasive technology (Fogg 2003), this kind of structuralization is intentional and may be justified for some purposes. However, implicitly we thereby adopt and accept before-the-fact assumptions made by systems designers about operational conditions, desired effects, appropriate routes of progress, and behaviors. As I see it, an undesired consequence of technological structuralization is that the behavior of individuals and system effects in real-life situations then become governed by model-based predictions and predetermined values, instead of being behaviors and system effects locally controlled to suit situated aspects and contextually meaningful values. When effects are evaluated and values are defined from a *detached* perspective (e.g., before the fact), the richness and meaning of the *involved* perspective is lost (Dreyfus 1972, 1986, 1992, Dreyfus and Dreyfus 1988). While detached findings might be specific and rigorous thereby appearing thoroughly convincing, they may still have no meaning because they lack real-life relevance (Boulding 1956, Benbasat and Zmud 1999, Davenport and Markus 1999), which is a situation leading to ethical dilemmas when designing and implementing persuasive technologies (Fogg 2003, chap. 9). The lost involved perspective and technological systems that counteract the gaining of edge awareness have, arguably, severe consequences.

For scrutiny of this problem, a number of assumptions will be challenged, connecting counteracted edge awareness to interrelated problems. I will, for example, argue that means for involved and situated (i.e., in the loop) system control are essential for edge awareness, for rich contextual interpretations of system effects, and for local judgments of values. Without situated controllability, the local nature of a system becomes obscured, implying that the possibility of locally adjusted system effects and contextually relevant values will remain unknown. The lack of locally relevant values is a consequence that from a detached perspective is a non-existing issue as it is not part of the equation. Technological usefulness, when defined by model-based aspects, becomes thereby in effect a self-fulfilling prophecy because there are no means to discover alternative values. Model-based evaluations and formal definitions of values tend to harmonize with model-based scenarios governing system designs because they are all detached descriptions thereby similar in character.

Out-of-the-loop performance problems are associated with several complex issues (Dekker and Woods 2002, Dekker and Hollnagel 2004, Parasuraman *et al.* 2008) such as vigilance decrements, complacency and over-trust in automation, as well as control skill decay (Endsley and Kiris 1995, Kaber and Endsley 1997, Endsley and Kaber 1999). Moreover, a lack of edge awareness obstructs the developing of incentives, for example to diverge consciously from predetermined routes, does it not? I will argue that

this is actually the case because, without sufficient edge awareness, options and consequences become difficult to discern. The predetermined route becomes then the only route. To consciously diverge from a predefined route is, however, analogous to rejecting a proposal or withstand persuasion, an essential aspect of autonomous thinking and of having a free will. Edge awareness is therefore argued as being crucial for human emancipation and *autonomy* because to adopt and accept externally defined values and behaviors implies *heteronomy* (Kant 1785, Stensson and Jansson 2014a).

The difference between model-based and situated usefulness is somewhat similar to that between theory and practice, or between knowledge and skill. While detached notions of usefulness may be theoretically valid, in practice, such notions might be largely irrelevant, and while knowledge often is a prerequisite for success, without skills to go through with activities known to be effective they will probably fail. Hence, model-based aspects are perhaps necessary, but they appear not sufficient for achieving true usefulness for technology.

Within a calculative safety culture, focus is on predictability (Reason 1998, Westrum 2004a, Parker *et al.* 2006), a strategy that at first may give the impression of actually achieving safety. This impression is, however, a *safety paradox* because efforts to enforce predictability may in fact cause disasters in a dynamic world requiring the kind of variability that comes from human subjectivity and creativity, a variability required for an ability to discover hazards and adjust to local conditions in order to preserve safety (Reason 2000, 2008, Vicente 2006). It is a paradox because enforced predictability (i.e., tight coupling) in complex environments makes accidents and disasters become normal consequences of systems design (Perrow 1999). Phrased analogously for usefulness as for safety, within a calculative usefulness culture, focus on predetermined desires make systems irrelevant because situated and subjective adaptations are required for achieving contextually relevant effects. It will here be argued that, besides normal accidents, a calculative design culture implies normal uselessness as well.

The identified problem is that the calculative model-based approach deliberately avoids the subjective aspects and value relations that are what determine whether systems are considered useful. Moreover, the subjective considering of a system as useful is, arguably, a necessary condition for a system to actually be used creatively thereby having the potential to become locally useful, as when participating in making unpredicted hazardous situations safe. Whether to focus on predetermination or situated control is a matter of choice and the present paradigm seems to have a strong calculative preference. It seems the norm still is that of industrial age technical rationality, which is a paradigm that Schön (1983) describes as the heritage of positivism and the result of an insistent and purposeful cleansing of subjectivity from knowledge and moral matters that has been going on since the reformation. For technical rationality to make sense, the detached

perspective is required. Furthermore, because of history and inveterate habit, the detached perspective seems today more or less implicitly assumed whenever striving for scientific rigor, which in turn makes situated and practical aspects more or less deliberately disregarded. The fear of uncertainty has also become a flight from experience (Reed 1996, chap. 3), showing as that detached predictability often is regarded more valuable than involved abilities. Implicitly, the flight from experience becomes also a flight from responsibility because objective predictability removes the need for responsible subjective decisions.

This research is critical, relying on insight, critique, and transformative redefinition (Myers and Klein 2011), but it is not critical towards the present paradigm of technical rationality as an end in itself. Research questions have been generated through problematization and a challenging of underlying assumptions (Alvesson and Sandberg 2011, Sandberg and Alvesson 2011), resulting in identification of certain deficiencies in the present paradigm considered necessary to critique. Because the challenged paradigm appears rather firmly rooted in current society, in particular in some highly influential areas such as high-tech industry and government (including the military), the contributed alternative theories should at least be interesting as a rhetoric statement (Davis 1971, 1986). This research relies on *critical realism* (Bhaskar 2008, Collier 1994, Archer *et al.* 2007) as the underlying philosophy of science and thereby on the transformational model of social activity (Faulkner and Runde 2013). For critical realist science, the primary goal is to explain studied phenomena and redefine them based on explanations conceptualized by *retroduction* to plausible structures and underlying mechanisms (Wynn and Williams 2012, Mingers *et al.* 2013). Such underlying structures explaining usefulness must obviously include safety because a too dangerous system can hardly be useful, and for the concept of safety, a similar discrepancy as that between the contemporary view of usefulness and the contributed redefinition has apparently been identified, resulting in the complementary concept of resilience (Hollnagel *et al.* 2006, Sheridan 2008). Usefulness does, however, besides safety and resilience also require effectiveness or efficacy, for which a concept corresponding to resilience seems missing. The contributed redefinition of usefulness fills this sub-structural gap by suggesting tentatively *potentiality* as the involved concept, complementary to the more detached concept of effectiveness. The present redefinition of usefulness connects these four concepts (effectiveness, safety, potentiality, and resilience) in a model explaining situated usefulness.

Besides these four concepts, to explain situated technological usefulness, human awareness of what is to be achieved in a particular situation and how to achieve this with the technology at hand appear essential. Within the area of human factors, perhaps particularly in the field of aviation, the concept of situation awareness has been influential from as far back as the First World

War (Endsley 1995a, 1995b, Wickens 2008). The concept, however, has always been slightly controversial and somewhat problematic (e.g., Sarter and Woods 1991, Flach 1995, Salmon *et al.* 2008, Stanton *et al.* 2010). For example, what in a situation is the awareness about? One definition of situation awareness often used, contributed by Endsley (1995b), suggests that the awareness is about elements in the environment. However, the question remains, what is it an environment of? Is it the immediate environment around the human being having the awareness, or is it the environment of the user together with the technological system of concern? Furthermore, is situation awareness residing in-mind as the psychological approach identified with Endsley suggests, in-world as the engineering view maintains, or in-interaction as systems ergonomics with a sociotechnical systems approach hold (Stanton *et al.* 2010)? While the present view aligns somewhat with the sociotechnical systems approach, the purpose of this research requires special focus on the interaction between technology and system-controlling individuals, making psychological aspects significant.

The contributed redefinition of usefulness takes stance in the view that regardless the level of automation, situation awareness, or lack thereof, for human system operators (i.e., users) is significant for what effects that actually will occur in real-life situations. With a fully automated system, the operator has marginal influence and effects become governed by before-the-fact considerations and designed system properties. Situation awareness does not matter much when using highly automated systems, a fact therefore suitable as an indicator of calculative usefulness. For a more manually controlled system, effects are to a greater extent the result of how system properties are applied in the situation by the operator, thus depending on how the operator interpret the specific situation, on the actual capabilities and properties of the system, and on the operator's ability to control the system. The discrepancy between these approaches corresponds in a sense to the distinction between the classic view of human-decision making and dynamic or naturalistic decision-making (e.g., Brehmer 1992, Cannon-Bowers *et al.* 1996, Klein 2008). Situation awareness seems intuitively relevant for all aspects of manual control, thereby indicative for aspects of situated usefulness. Explanations of underlying structures and mechanisms for situation awareness, or lack thereof, can therefore help explain situated usefulness and, presumably, guide the design of truly useful technology.

The case of manual control provides thereby the frame of reference for another contribution of this work. In order to reduce the vagueness of the concept, or of some of its interpretations, situation awareness is here complemented by the concepts of *system awareness* and *edge awareness*. System awareness, meaning awareness about the technological system of concern, is explicitly designating the technological system as distinct from both the user and the environment, while simultaneously depicting it as a crucial element of which to be aware. Furthermore, the fact that many

sought-after effects occur from the interaction between the technological system and the environment makes awareness about interaction effects essential as well. Presumably, awareness about interaction effects will benefit from explicating properties of interacting entities (i.e., the technology and its environment) and interaction-effects-awareness seems thereby relevant for the ability to control manually the technology in a situated manner. Interaction relations such as performance characteristics tend to include performance edges, of which friction normally is a familiar example. Friction is an interaction relation that leads to slipping if the performance edge is crossed (cf. ch. 1.4.2 as well as Figure 9.2 in ch. 9.3, p. 154).

With these complementary awareness concepts, the case of automated control can get a richer description. While the operator may have good situation awareness, automation tend, often ironically (Bainbridge 1983), to imply bad system awareness (Endsley and Kaber 1999), for example when operators are suffering from mode confusion thereby unaware of actual system workings (Sarter and Woods 1995). Without appropriate system awareness, situated system-environment interaction characteristics cannot be comprehended sufficiently well, and bad edge awareness becomes a direct consequence, and situation awareness of low significance. Automated technology will thereby maintain and enforce its calculative usefulness as long as conditions align sufficiently with what was predicted. However, if conditions become overly unfavorable, system performance edges will be crossed without operator insurrection, despite possibly sufficient situation awareness simply because the edge awareness is insufficient. For such situations, the model-based usefulness is perhaps satisfactory, but the *situated usefulness* is evidently inadequate. Hence, aiming to design technological systems facilitating situated usefulness implies the title phrase, *a quest for edge awareness*.

1.2 Research objectives

Human beings must necessarily be involved somehow for situations to really matter for human beings, an obvious truism, no doubt. Nevertheless, people are key players in situations that matter because values determining what matters and judgments of what is desired are subjective to people associated with a situation. The relation to some involved human beings might be distant and indirect (e.g., for remotely operated vehicles), a fact not implying matters of less value automatically, rather it tends to imply relations more complicated to analyze. The presence of technology is, on the other hand, optional, although, in our modern world it tends to be the case. Technology is today often either a key factor to what happens, a necessary condition for certain things to happen, or by own merits momentous enough to make the situation matter (i.e., expensive, exclusive, impressive, menacing, etc.). The

relation between entities, system components, agents, players, etc.,[2] is clearly quite complex. This relation between technological systems and human beings is the concern of the present research.

The overall purpose is to explore usefulness in order to follow up on an initial sensing of a disproportion within the contemporary view of usefulness, strive towards further and richer insights, and contribute a well-grounded critique of the prevailing view. The main objective is to present an alternative definition of usefulness that aims to transform the contemporary view into something more beneficial for humankind.

Research purpose: The purpose is to scrutinize and reflect on the contemporary view of the concept of usefulness. This means to explore usefulness in order to find out what makes certain systems useful, also in situations where presumed conditions for using the technology of concern fails to be perfectly satisfied, or when predicted purposes of use are irrelevant. It means to find out what prevents some systems from being useful despite having similar performance properties compared to systems actually considered useful. The purpose is to contrast *model-based* usefulness against *situated* usefulness.

A **second purpose** is to reflect critically on the impact that the prevailing view of technological usefulness has on human autonomy. For this purpose, it is necessary to stress that the meaning of certain concepts related to the use of technology affects how people view their role in relation to the technology they use. Usefulness is such a concept. It is a concept that intuitively connects situated human activities with values of results. However, the contemporary reduction of usefulness to be defined merely by detached model-based values risk making the involved perspective and the value of involved autonomous human thinking obsolete.

Research objective: The objective is to contribute an alternative definition of usefulness that takes into account also the situated and the contextual as well as the unpredictable. The purpose of contributing a redefinition is to put more focus on the involved perspective and on local values, and thereby initiate a social change.

1.2.1 General research setting

For initially sensing the suggested disproportion within the contemporary view of usefulness, my professional background seems relevant as it is of a kind where situated usefulness of high-performance technological systems tends to become a matter of life and death. Appropriate usefulness is then not

[2]This is a difficult choice of word. To call human beings components or entities may give a too mechanistic impression and to designate technological systems as agents or players might give an impression of conscious deliberation not applicable to artifacts.

a luxury as it might be for non-critical everyday gadgets, and usefulness-inhibiting system designs are then more problematic than merely annoying. Dealing with this kind of usefulness is arguably adding some gravity to the research questions. Presumably, the rather unusual structure of my earlier career is also relevant for identification of plausible reasons for the suggested disproportion (i.e., relevant for retroduction). My background connects also to the research approach (presented in ch. 2), as well as to the choice of explored usefulness-related concepts, thereby shaping the overall structure of this thesis (outlined in ch. 1.2.3). Moreover, knowledge and experiences from my former line of work constitute a significant set of preconceptions undoubtedly influential on the present research. For all the above reasons, a narrated description of selected aspects of my background is provided in chapter 3, but for the present framing of research, a short outline seems required as well.

Around twenty-five years ago, I passed the screening process, was accepted into flight-school, managed to complete all training stages, graduated as an officer, and became a Swedish Air Force fighter pilot. About five years later, I applied to extend my education to become a pilot engineer,[3] which made me shift gradually from working full-time at a squadron to focus more on technology development, training and education, and military meta-scientific[4] studies, while continue flying fighters until the type I was trained on (AJ/S/JA37 Viggen) retired from service. Experience from professional handling of fighter aircraft in military contexts appears valuable when the phenomenon studied is usefulness of high-performance technological systems in uncertain situations. In fact, the Human Factors (HF) research discipline, my primary area of interest as a pilot engineer, originates from the field of military aviation.[5]

The initial primary focus on the case of use (i.e., on situations in which systems actually are used, as when flying) seems in light of the above rather natural. However, the properties that may or may not make a certain technological system useful in the case of use are largely determined in advance, by its design. This before-the-case determination is arguably true for all kinds of artificial systems including adaptive and self-modifying ones because the limits for their adaptation and the principles for their self-modification are determined by design decisions. Both explicit and implicit (e.g., subconscious) assumptions made by system designers about the reality

[3]A pilot engineer is (in Sweden) a flying officer with a M.Sc. degree, in my case a degree in engineering physics obtained at Uppsala University, specializing in computer systems.

[4]This notion should not be interpreted as a condescending one. On the contrary, I maintain that such 'meta-scientific' research projects often represent a good balance between rigor and relevance. More about this in chapter 3, particularly in chapter 3.3.

[5]The origin of the modern Human Factors and Ergonomics research discipline is often associated with the establishment of two aeronautical research labs in the early 1930s, just after World War I, one at Brooks Air Force Base and one at Wright-Patterson Air Force Base, both in USA (e.g., http://en.wikipedia.org/wiki/Human_factors_and_ergonomics).

in which the system is to be used and about how and why the system is to be used in that particular reality become implemented as a kind of hard-coded and system built-in prejudice. These presuppositions affect performance properties and the character of system control in the case of use. Hence, the case of design is necessary to include when exploring usefulness.

1.2.2 Research questions

This thesis is divided into three phases: theory generation, case studies, and discussions. The first phase is governed by the pursuing of RQ1, a question focusing on exploration of *the concept* of usefulness, resulting in a set of conceptual frameworks and models or typologies identified as relevant for describing situated usefulness.

The theoretical contribution from phase one is then used in phase two for scrutiny of four real-world cases from two perspectives. RQ2 focuses on the involved perspective in the case of use and RQ3 on the detached perspective in the case of design. In conjunction, the results from pursuing the three research questions provide the grounds on which the alternative definition of usefulness is based, the research objective. The second research purpose, to reflect on the impact of the prevailing view of usefulness and the counterproductive shift in concept meanings, is addressed in phase three.

(RQ1) **The concept of usefulness**:
What is the nature of *situated usefulness*? What makes a technological system useful, even when it is used in not completely predictable real-life situations?

(RQ2) **The case of use**:
What is the role of a human user when using potentially useful technology and how does this role affect actual usefulness?

(RQ3) **The case of design**:
In the case of design, how does the view of the human role in the case of use affect the usefulness of the designed technology?

1.2.3 Structure of thesis

The thesis is structured in four major sections, illustrated by Figure 1.1. The first section is the preface that besides this introduction consists of a more detailed presentation of the actual research approach, including descriptions of philosophical and methodological assumptions (ch. 2). The preface ends with a narrative of selected aspects of my background (ch. 3), structured in a manner corresponding with how the theoretical contribution generated in phase one is presented. The remaining introduction (ch. 1), aims to further

outline the research topic and provide a richer frame of reference by taking this rather complex research topic, a topic concerning the abstract, dependent on subjective values, and socially situated concept of usefulness, and relate it to concrete everyday examples such as car driving and airline traveling. In addition, a 'cartoonish' metaphor of the title concept, edge awareness, is provided. Some examples are returned to later in the thesis.

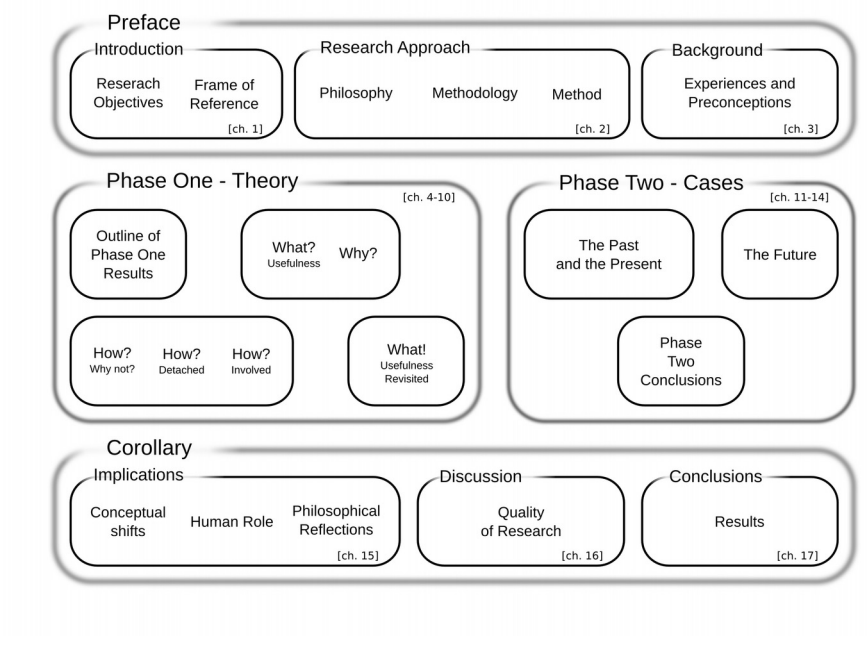


Figure 1.1: Thesis structure

The three following sections represent the three research phases. The first phase, theory generation, is presented in chapters 4 to 10. Because the purpose of this research is to scrutinize the contemporary view of a concept, novel frameworks, models, and typologies, are required for analysis and for the ability to convey alternative meanings. The concept of usefulness is scrutinized by problematization, initially by pursuing RQ1 – what is the nature of usefulness? It is, however, pointless to talk about usefulness without knowing *what* should be useful (what thing), *for what purpose* and *for whom*, *when* and *where* it would be useful to use this thing (the what), and *how* this thing is to become useful in real-life situations. By asking how usefulness comes to be in the case of use and how the case of use comes to be from the case of design, a connection is established between the three research questions. For clarity, chapter 4 – Outline of phase one results, opens up phase one with an overview of the theoretical contribution.

In chapter 5 – Exploring usefulness, a worldview is presented that explicates relevant environments thereby providing means for describing

where, when, and for whom the technology should be useful. The worldview outlines the technological system of concern, and frameworks for describing characteristic properties of the technology (i.e., the what) are presented. Chapter 6 – Desired effects, addresses the question for what purpose (i.e., why) the system should be useful, by introducing a *tension* named the character of utility, which is a tension between possibly contradicting aspects relevant for achieving desired effects and for avoiding undesired effects.

Beginning with chapter 7 – Uncertainty and unpredictability, focus turns towards how the technological system of concern can become useful within the actual situation. First, plausible reasons are explored for why there are problems with knowing in advance whether a system will be useful. A framework for describing the character of uncertainty is presented and a road map (i.e., a model) is introduced, showing the road from conditions to effects, in theory. Another tension is introduced as well, initially with two levels, first, the approach to controllability, the case of design, and second, the character of controllability, the design result. Chapter 8 – Models and scenarios, and simulators, explores aspects affecting the human role in relation to technological systems, it explores the human contribution to system effects. The road map is then updated with aspects of how to get from conditions to effects, in practice. The tension within controllability is also updated with a third level, denoted the character of control, the case of use. In chapter 9 – Situation awareness, focus is on the human component. The exploration of the human being as a system controller is made by scrutinizing situation awareness, the influential but difficult concept often associated with human factors in aviation. Situation awareness is complemented with system awareness and with the thesis title concept of edge awareness, three concepts that in conjunction depict significant aspects for situated usefulness.

Chapter 10 – Explaining usefulness, is directly addressing the research objective and RQ1. A conceptual framework is presented, aiming to link together all the concepts, models, frameworks, and typologies, from chapters 5 to 9. The alternative definition of usefulness is then presented as a two-by-two matrix describing core aspects of the character of situated usefulness. The two dimensions of the matrix are the two tensions presented in chapters 6.2 and 8.2, the tension between achieving desired effects and avoiding undesired effects (i.e., the character of utility), and the tension between model based and situated controllability (i.e., the character of controllability). The latter tension is at the core for pursuing RQ2 and RQ3 in the empirical phase two.

Figure 1.2: SAS SK751 wreckage at Gottröra, a crash caused by ice breaking off from the wings damaging the engines, or, is inappropriate human interaction with automated systems a more plausible explanation? (Photo: SHK)

The second phase, the empirical studies, is presented in chapters 11 to 14. Four major cases plus three future scenarios are scrutinized by applying the theoretical contribution generated in phase one. Alternative qualitative interpretations of the events are then made according to the methodology described in chapter 2.3.2. The four cases are all accidents and disasters, which is in accordance with the methodological primacy of the pathological; "By seeing how something goes wrong we find out more about the conditions of its working properly than we ever would by observing it working properly" (Collier 1994, p. 165). The cases studied are two airliner and two nuclear power plant incidents: First, chapter 12.1 – "The Christmas miracle" in Gottröra (Figure 1.2) on December 27th,1991, when SAS SK751 had both engines destroyed shortly after take-off and made an emergency landing without casualties (although a few were severely injured). Second, chapter 12.2 – An incident at the Forsmark nuclear power station on July 25th, 2006, when external power was lost and the uninterruptible power supply system did not work as expected (but in the end without any effects on the surroundings). Third, chapter 12.3 – Air France AF447 that disappeared in the Atlantic on June 1st, 2009, with 228 people on-board. Fourth, chapter 12.4 – The Fukushima Daiichi disaster, beginning with The Great Eastern Earthquake on March 11th, 2011.

For the future scenarios, this methodological stance of studying the pathological is, however, practically impossible, simply because developers

of future technologies tend to focus on providing accounts of plausible success scenarios, not failures. The three examples of future technological systems are therefore treated as one additional case of slightly lower empirical significance, mostly as 'food for thought'. However, having some kind of account for the future is thoroughly interesting when addressing RQ3, the case of design.

The third phase, the corollary, is presented in chapters 15 to 17, consisting of a discussion about possible implications of this research. The discussion addresses also the second objective, to reflect on the impact from the contemporary view of usefulness on human autonomy. Finally, the thesis ends with a summary and scrutiny of research findings.

1.3 Setting the frame of reference

1.3.1 Edges are everywhere

One problem with edges is that for any real-life situation, reality will most certainly bring more than one edge necessary to be aware of, and presumably, many of these edges will have completely different characteristics. Sometimes edges can be sharp and require immediate and skillful control of edge-related activities, such as when suddenly getting a flat tire at highway speed in snowy conditions instantly bringing your car to the verge of skidding. Other edges are smooth and develop slowly over time thus possible to 'get the hang of' before things become ugly, such as when shaping for a strain injury by performing continuously certain physical labor badly. If not identified and addressed by adjusting your behavior, eventually you will fall over the edge and make yourself suffer from permanent damage. Edges must not, however, be technological or physical as in these two examples, they may also be psychological and social, and abstract, such as rules and ideas that nevertheless can make a situation truly unfavorable if not complied with, for example when triggering a riot by showing unaccepted behavior on social media. Social and psychological edges are in addition often thoroughly intertwined. The failure to realize that something is considered crucial in a certain context can be thought of as failing to identify and thereby falling over the social edge of trust and reliability. At the same time, the failure to realize that a certain aspect is crucial can be thought of as failing to identify and thereby falling over the psychological edge of responsibility. The taking of responsibility and the gaining of trust are two sides of the same coin.

Another problem is that for any real-life situation, reality will most certainly bring several edges necessary to be aware of simultaneously, and presumably, it will be a mixture of physical and psychological as well as

social edges. We live in a social world populated by autonomous human beings inherently predisposed to attribute meaning to things, a world in which local conditions and different backgrounds inescapably lead to varying interpretations. We are human beings capable of coming up with new ideas potentially opposing old ones and thereby clash with established norms, which arguably is a sign of health and a prerequisite for autonomy and social evolution. However, personal ideas and differently interpreted meanings will lead to competing interests that in turn make social edges incompatible with each other because they depend on incomparable factors. The world is, in addition, a world in which people tend to make use of technology in their purposeful activities. The modern world may therefore be described as a sociotechnical system (Trist and Bamforth 1951, Trist 1981), a system that combines and intertwines reasonably well known and generally valid physical or technological factors with subjectively defined and contextually dependent psychological or social factors. One direct consequence of this intertwining of factors is that while physical factors might righteously be considered practically indisputable, what indisputable factors that are determined important in a specific situation are by all means disputable, and in addition, thoroughly dependent on the social context.

Maneuvering through a world sporting such a dynamic maze of interdependent but often incompatible edges implies highly complex decision situations that call for careful judgments. The judgments must be careful because it is unlikely that decision situations in an open sociotechnical system will allow for anyone in advance identifiable alternative of action to be known as unquestionably correct. How can an objectively correct action possibly exist among subjectively defined options? Sociotechnical systems are inherently open, as all system-views are in practice, simply because any system-view describing a problem situation is subjective and thereby susceptible to critique, meaning that there might be another system-view describing the problem better. The inescapable openness implies that the system is non-deterministic for any chosen system-view, making the idea of a predetermined objectively correct solution irrelevant. Perhaps, but arguably not without considerable analytical effort, it may be possible with hindsight to agree on what would have been objectively correct to do in a previous and well-documented situation, although this would in some sense also be slightly irrelevant since the decision has already been made. What still motivates after-the-fact analytical efforts is to gain lessons-learned knowledge potentially useful in future situations resembling the analyzed one. Nevertheless, without the possibility of a before the fact knowable objectively correct option, in practice, most decisions come down to pragmatics and what feels right and makes sense at the moment (Rachels 2009). Such decisions are ideally based on a healthy amount of common sense and a sufficiently well developed 'gut-feeling',

although such rather vague human qualities presuppose, naturally, a reassuring amount of knowledge about things actually knowable in advance.

There is, however, yet another problem, which is the main concern of this thesis. It is the fact that modern life tend to include an astounding and increasing amount of more and more advanced technological systems that interfere with many real-life edges. Such interference is today ubiquitous because complex technologies are everywhere and occupy all conceivable roles from being non-critical nice-to-have gadgets to being fundamental need-to-have means required for certain activities. Technological means are for example required when controlling delicate processes such as those going on in nuclear power plants, or when flying aerodynamically unstable aircraft. The ubiquity of these advanced technologies make, however, their designs obtrusively influential and today technological systems tend to be quite extensively computerized, a fact that comes with a risk of making already difficult decision situations even more problematic. This development is a real problem because computerization tends to obscure natural and intuitively understandable edges, and introduce new edges of a kind that seems inherently difficult for human beings to handle. Utilizing the powerful design possibilities that computers bring risk thereby, without careful considerations, oppose the development of relevant gut-feelings required for human beings to make insightful situated decisions. The antidote is, supposedly, to explicitly design for edge awareness in order to facilitate situated usefulness.

1.3.2 Usefulness, of what?

The present work is all about usefulness and relevance of advanced, often computerized, and highly integrated, technological systems of systems, which perhaps is a notion that includes almost all technological things today. Focus is, however, primarily on distinguishable 'gadgets', for the ability to discern the technological objects of concern. Particularly, focus is on vehicles and other real-time systems, for the ability to discern directly issues and effects. Furthermore, focus is on usage of such technologies in situations that are unpredictable in some sense, for the ability to discriminate truly useful technological properties from self-fulfilling prophecies of artificial usefulness (i.e., to distinguish between situated and model-based usefulness). The latter occurs when the world effectively is forced to comply with assumptions built into technological designs in order to justify system behaviors. My view is that most real-world situations are more unpredictable than we might want them to be, or perhaps choose to regard them as. Especially if situation as a concept includes context, purpose, and other social and psychological aspects subject to local variations and competing interests. This specialized focus on real-time systems in hostile environments does not mean that normal information systems (often intertwined with

organizational processes and distributed throughout the society over the Internet) in friendly environments or traditional stand-alone office-like personal computer systems with well-defined purposes are disregarded. My view is that what is discussed in this thesis should be relevant for all kinds of technology. There is, however, a crucial difference between information systems or computerized personal gadgets and systems like fighter aircraft. Undesired consequences of and issues related to, say, the implementation of a corporate information system may have to evolve for years before they can be properly identified, whereas undesired effects of principally identical technological properties tend to become evident rather quickly in a fighter jet at the speed of sound. When dodging intelligent adversaries actively trying to exploit possible weaknesses of the technological system you are using, situated usefulness can often be more valuable than model-based high performance measures. To scrutinize aspects by pushing them to the extreme like in this fighter aircraft example is both a rhetoric theme of the present work and a way to bring issues, perhaps neglected because under normal conditions they are indirect and somewhat subtle, out in bright light.

1.4 Advanced technology, for better or worse?

This thesis is not a crusade against technological advances and computers, although some passages might give that impression, perhaps because of the rhetoric approach of taking questions to the extreme. On the contrary, I have a sincere interest in technology and I am mostly quite positive towards advanced technological systems, and this is perhaps particularly true when it comes to computers in general.

However, I have come to regard current practice as severely afflicted with an unwarranted appraisal of model-consistent computer-like behavior, for technological systems such as machines, naturally, but unfortunately also for social systems and human beings. The contemporary view of usefulness, as well as that of safety, are examples of such a calculative view, a view based on an exaggerated belief in the attainability of pure objectivity and in that it actually exist unquestionably correct courses of actions knowable in advance. In particular, it appears that a strong belief based on this objectivity ideal prevails, a belief in the attainability of absolute safety. With this view, everything would be safe as in completely predictable, if only it was possible to do away with, or harness, all distracting open factors to otherwise closed system models and elaborate scenarios, all distracting factors such as irrational human emotions and personal ideas. Nothing unpredictable and dangerous would happen, if only it was possible to make people start following such presumed objectively correct courses of actions. The grounding assumption seems to be that the irrational human behavior is the culprit, that autonomous human actions are causing dangerous

unpredictability in an otherwise orderly reality with essentially predictable outcomes by use of scientifically proven state-of-the-art models. Since Galileo, who allegedly coined the phrase, objective mathematics has been considered 'the pure language of science', which, however, is a language strong on figures and formal relations but weak on values and subjective contextual meanings. Without human-induced unpredictability, it appears the optimal behavior for every conceivable situation is considered derivable, as well as possible to derive in advance, if only there was a computational power capable of maintaining sufficiently elaborate models as well as capable of analyzing these models exhaustively. This, I believe, is to try actively to revive the old Daemon of Laplace.[6] Luckily, as I see it, and despite the firm grip this predictability-addiction seems to have on modern society, the whole idea must clearly be a misconception because full predictability is an irrelevant idea for anything but trivial and completely closed systems describing problems of simplicity (Weaver 1948). For problem situations that include social aspects, for social systems that comprise wicked problems (Churchman 1967a), there is no such thing as a definitive and objective answer (Rittel and Webber 1973).

It seems that fifty years ago or so, the idea of objectively correct behavior based on systemic analysis and calculable system models was not as highly regarded as today. Are we perhaps a bit dazzled by the impressive properties of modern computers, making us unable to see clearly what computers and models lack? While the computational powers of computers undoubtedly continue to conquer new grounds facilitating beneficial new kinds of technological designs, is not the principle discrepancy identified long ago between general models and true reality still valid? Today, however, the exaggerated faith in models shows as continuously increasing efforts to enforce predetermined system workings presupposed to be the correct way of doing things, and the enforcement is often facilitated by technological means such as computerized automations.

What seems to be forgotten is that these predetermined system behaviors are based on system models that inescapably are incomplete and simplified representations of reality. However, “All models are wrong” (Box 1976, p. 792, Sterman 2002). What also seems to be forgotten is that these models are always presupposing certain conditions and assuming certain intentions and goals for activities, intentions and goals that perhaps are not perfectly relevant in situations where the systems later participate. Technological structuralization comes from being governed by such models, resulting in an imposition of a stereotypical and context-insensitive order on human activities, an order that ultimately occur at the expense of human autonomy. What at first appears merely curious, but becomes frightening when thinking

[6]For more information about Laplace's Daemon, please consult for example the web: e.g., http://en.wikipedia.org/wiki/Laplace's_demon

more about it, is that often the intention with such a stereotypically behaving system actually is to support human beings. The scary part is that when people later interfere with the system because they consider the support irrelevant or inappropriate, the norm is to vouch for the stereotypical behavior of the technology. The mathematically derived behavior is judged correct because it is based on so-called impeccable models freed from personal bias and irrational human desires. This calculative approach clearly shows that the paradigm of technical rationality prevails, despite its shortcomings. For safeguard systems, a safety paradox proves the calculative approach counterproductive (Reason 2000). The present research suggests that there is an analogous usefulness paradox.

Although, computerized systems do have some useful properties, at least potentially useful. The problem is that systems tend to be implemented in a way that inhibits people from making use of these properties such that they become truly useful in real situations. There seems to be an unhealthy balance of design efforts, where the exaggerated focus on model-based usefulness and calculative safety implies a disregarding of aspects necessary for situated usefulness and resilient safety. That is, this thesis should be viewed as an argument against uncritical acceptance of technological influence based on simplified and stereotypical models commonly implemented in software, rather than as a crusade against computerized technology and automation. Misuse of technology cannot be blamed on technology, only on people designing it and on people choosing to use it. Responsibility of actions is therefore a key issue, and because only human beings can fully understand human values, only human beings can be responsible for actions affecting human beings.

1.4.1 The categorical limit for technology

The problem with paradoxical or counterproductive aspects of otherwise beneficial new technologies is that these kinds of issues are often thoroughly difficult to identify because they tend to be implicit or appear only as long-term consequences. Such undesired consequences are in some sense the murky shadows of our shiny new gadgets, and we seem currently to choose, more or less consciously, to disregard the unwanted. I am (mostly) all for technological progress, but I am also all for human emancipation, and I regard human beings unquestionably superior to technology when it comes to authority and autonomy, regardless the so-called weaknesses of the human mind. These weaknesses center on our inability to keep emotions and subjectivity away from interfering with our decisions, but arguably, that is not a problem, it is what makes us human. The idea of perfect objectivity as the ideal for human activities is, I believe, completely wrong. It is a delusion because our emotions and subjectivity are what gives meaning to our actions. Emotions and subjectivity are perhaps the two most important

aspects that distinguish human beings from technology and they are in the end what merit our autonomy, making the contemporary hype around autonomous technological systems a delusion as well. In fact, technology cannot be autonomous, despite this conception apparently being a contemporary buzzword among technologists and systems manufacturers (elaborated on in ch. 15.1). My view is that the reason why the designation of technological systems as autonomous in some areas is considered appropriate is because there is an ongoing and from a human emancipation perspective counterproductive reduction of concept richness, for several of the concepts discussed throughout this thesis. Autonomy and safety are but two examples, and usefulness is the main concern. The contemporary use of autonomy among technologists is arguably to forget some aspects and confuse other crucial aspects of the concept. In particular, what seems to be forgotten is that being autonomous is about having the right to make subjectively biased decisions, a right justified for human beings by the fact that they stand responsible, ultimately by putting their lives at stake, for every decision they make. Technological systems have no life and can therefore never stand responsible for their actions and thereby never be autonomous, a categorical limit for technology (ch. 15.2).

1.4.2 Automated safety and skidding cars

Regardless wherever the categorical limit for automation is drawn, the present position is that technology can and should be designed to help, guide, guard, and inform human beings, but never enforce control on human activities without proper means for people to intervene, which might be more problematic than it at first appears. It is problematic because to have proper means to intervene requires, arguably, also having proper means to develop the incentive to intervene, which in turn requires sufficient knowledge about system workings to know when and why to intervene, as well as sufficient skills to be confident about how to intervene. Without such means, people will not intervene unless overwhelmed by undesired effects. However, predetermined effects are not necessarily the same effects as what involved human beings would want to create, if they could be aware of what effects the system might be capable of creating, and if they were aware of how to work the system comprehensively, if they had the skills to do it. The prophecy of technological success for automations becomes self-fulfilling because people are essentially forced to align with predetermined goals when alternative goals are obscured by system designs. Systems must be useful if people use them according to their design goals, must they not? The usefulness paradox occurs when the seductive powers of technological advances enforce a calculative approach that oppose designing means for situated system control the and gaining of knowledge about system workings in the case of use, thereby reducing practical usefulness. My view is that this

paradox is real, but hopefully manageable, if sufficient reflection and scrutiny is applied before systems are designed. The important question is how we make use of technological advancements. What I wonder is why we seem to use them so often to reduce our own influence in situations where our lives are at stake?

Consider for instance the following safety dilemma. Few would say that technological safety systems in general are for the worse, neither would most people, if any, say that human virtues such as experience knowledge and skill are undesired qualities. What about system designs exploiting technological advances such that they obstruct the gaining of these virtues, are they for good? Is it acceptable that technological designs counteract the gaining of experience knowledge and skill, if it is done in the name of safety? What about advanced technological measures intended to increase safety that in fact are insufficient and still leave to human operators the task to cope with remaining hazards, are such measures actually increasing safety? Has not the task to cope become even more difficult because of reduced experience knowledge and skill, thereby making it less probable for the human operator to succeed?

For example, if you always drive cars equipped with anti-lock braking systems, electronic stability control systems, and similar technologies intended to increase safety by imperceptibly assist you as the driver in keeping your vehicle from skidding, then it seems plausible that you, after a while, will become less skilled in recovering from skids. Perhaps even more important, you have probably also become less skilled in assessing the margins to the edge where your vehicle will start skidding. Compare this situation to the other approach, in which you are frequently collecting experience from consciously balancing on the edge of skidding, which inevitably also implies to occasionally having to recover from full-blown skids. Without the latter kind of experience and maneuvering skills, is it possible to be considered a truly safe driver? Arguably, skillful assessing of the margin to skidding is highly relevant also for several higher order driving decisions. Although some people might still assert that today, you do not really need skid-recovering skills because modern stability-control systems are better than most drivers ever can become, regardless how much they practice. Furthermore, people in general are actually unable to develop and maintain sufficient driving skills to master skids (perhaps only professional rally drivers can do this), but non-rally-professionals are still righteously allowed access to the traffic system. Hence, the traffic society as a whole appears to benefit from these kinds of technologies, in regard of reduced accident counts.[7] While there is no question about fewer accidents being a

[7]However, there is also the theory of risk homeostatis (Wilde 1998). Perhaps is the reduced accident count in part a marginal effect, resulting from continuous introduction of new and more capable technologies, better roads, and more structuralized behavior by increased regulations. If so, the long term effect may be that people adjust their behavior slowly and

good thing, there might in fact be more to safety than cold calculative numbers, the ability to make responsible higher-level decisions for instance.

Traffic safety is definitely a sensitive subject, but it is also a good case to discuss because most people can relate to it. Safety as such is in addition an obvious and fundamental aspect of usefulness, the central concept of this thesis. The problem when focusing solely on the actual number of accidents, which only is possible to collect after the fact, is that this is merely one aspect of safety. It is an important aspect, no doubt, but not the entire truth. To focus on accident numbers is here argued to be an example of the reduced concept richness (ch. 15.1) that appears to follow from regarding the world too much according to context-independent (mathematical) models while forgetting about the importance of situated meanings. For safety, this reduction of concept richness fosters the calculative safety culture (Reason 1998, Westrum 2004a, Parker *et al.* 2006) focusing on predictability and detached numbers collected after the fact. With a calculative approach, in which it is mathematically possible to derive generally correct behaviors, the assumption follows that local desires in general must be the same as global desires, an assumption justifying a technology-facilitated enforcement of globally desired behaviors on local situations. However, this approach is to make things too simple, arguably, slightly resembling the assumption that a person with one foot on burning coal and the other in a bucket of ice in general is feeling rather comfortable. To make the inadequacy of the statistical generalization assumption sufficiently clear, a ridiculous metaphor is provided, also sporting a 'cartoonish' version of edge awareness.

1.4.3 Safety on a plateau, from two perspectives

Imagine a mountaintop, an enormous cliff, shaped as a large flat plateau, on which there is a considerably large population. It is discovered that if people are blindfolded, significantly fewer are falling off the cliff, which obviously is a good thing. One plausible explanation for this highly desirable effect is that people are then discouraged from straying around exploring the edges of the plateau and thereby ending up unsafely close to the edge. Occasionally, however, one or two still fall over the edge somewhere because people simply have to move around regardless the blindfold, not least because it tends to become rather crowded at the center of the plateau. From a global perspective, the blindfolding is significantly increasing the safety on a number-of-accidents basis. This is a verified fact and the rationale for imposing the blindfolded order. On the other hand, from any individual point of view, for a citizen of this plateau-society walking around during the days, it is unavoidable to feel safer if allowed to drop the blindfold, which after all

eventually they will start challenging the edges, possibly without realizing how close they are, a development that then might restore accident counts to previous or higher levels.

seems quite reasonable. It must obviously be safer if one is able to see the edge and thereby have some means for actively avoiding it, yet it is an indisputable fact that fewer people in general fall over the edge if everyone is blindfolded. How can it be so?

The difference between these two aspects of safety is essential for this work and it appears to include some kind of situated edge awareness. On the one hand, there is what could be called a static and detached form of safety, a calculative kind of safety, easily scalable to entire populations, by use of objective mathematics, perhaps therefore the label calculative. This form of safety is about being safe in a general and therefore context independent sense, which is a kind of safety that in the plateau metaphor can be described by the measured average distance to the edge. Average distance is a perfectly valid measure because, obviously, if everyone keeps a reassuring distance to the edge no one will be falling off the cliff, in general. On the other hand, there is what could be called a dynamic and involved form of safety, a generative kind of safety focusing on specific situations and local conditions. This form of safety is about avoiding becoming unsafe and about becoming safe if, somehow, the situation has become undesirably unsafe. In the plateau metaphor, becoming safe is about the possibility to control your position in relation to the edge, which to begin with requires means for assessing the direction and distance to the edge. For the possibility to become safe in practice, you must also have the ability to use some situated means of control. Such situated control may be necessary for getting away from the edge, if you for whatever mysterious reasons have ended up too close. Being blindfolded may all right increase global safety in the form of detached calculative numbers, but it will definitely inhibit most of the generative safety and local involvement in becoming safe.

To focus merely on the detached aspect of safety becomes in terms of skidding cars the considering of a person never involved in a skidding incident a safer driver than a rally professional with a number of skid-offs on the record. The detached perspective makes the absence of skid-offs after the fact give the impression of safe driving, while the involved perspective might reveal that on the road this person is completely unaware of repeatedly being a hairbreadth away from skidding, which ought to be considered unsafe. The rally driver, on the other hand, is perhaps slightly overconfident and a bit reckless, and statistics prove this person unsafe, while on the road, slippery surfaces and tricky curves are handled with remarkable skill, which ought to be considered reassuringly safe. The former person appears from a detached perspective safer than the rally driver does, while the rally driver seems safer from the involved perspective. In terms of the plateau-society metaphor, the detached perspective gives that all individuals have the same risk of falling off the cliff, silently ignoring the relevance of personal intentions that make the true risk reach from almost a 100% for a suicide candidate to almost nil for the most cautious citizens. While reducing the

risk on a population level, the blindfolding does not necessarily reduce the risk for any specific individual. Perhaps a cautious one will become suicidal when forced to wear a blindfold.

If both these two forms of safety can be accepted, then I would argue that a plausible cornerstone of the four safety paradoxes that Reason (2000) talks about has been identified (the paradoxes are recapitulated in ch. 7.1). It has to do with a confusion of perspectives. The detached global and generic perspective does not give the same answers as an involved local and specific perspective. The local situation can be safer under precisely those conditions that make the global situation more unsafe, and vice versa, conditions that make the global situation safer can make a local situation unsafe. In the end, though, global aspects are never more than general conditions, not necessarily relevant in a specific situation. For a specific case, what make a difference are always local conditions. When enforcing too strongly calculative safety measures inherently focusing on global conditions, the safety becomes brittle (Hollnagel *et al.* 2006) and the probability for an accident induced by unforeseen local conditions to be fatal is increased because of reduced generative safety. The means to intervene when things in a local situation get out of hand, despite all precautions, have effectively been removed. Safety has then become a question of detached model validity instead of a generative quality.

1.4.4 The brittle safety with modern airliners

The tragic disaster with Air France flight 447 from Rio de Janeiro bound for Paris on June the 1st, 2009, is a telling example from reality (as opposed to the artificial plateau metaphor). The kind of aircraft in question, the Airbus A330, is a modern glass-cockpit[8] airliner with a computerized fly-by-wire[9] flight-control system. Most of the system-conveyed flight information and data about aircraft-status is presented to the pilots in a highly abstracted manner, some of it neatly combined into graphical representations on fancy color displays, for better and worse, arguably. During normal flight also the control of the aircraft is performed in a highly abstracted manner, by giving steering orders with the flight-control joystick to the flight-control system, orders about where to fly and not about how to fly there. This design implies that the pilots normally do not fly the aircraft in a traditional sense. Instead, they operate the flight-control system, which under normal circumstances accounts for a beneficial reduction of pilot workload and lower the risk for pilot slips and mishaps.

[8]The term 'glass-cockpit' refers to the use of computer screens and touch-buttons in cockpit layouts, as opposed to traditional (mechanical) dials and levers, it suggests implicitly an underlying system design relying extensively on computers.
[9]See footnote 82, p. 204, for a short description of the term 'fly-by-wire'.

The problem is that while accomplishing this simplified and clear-cut work environment, the design separates also the pilots mentally and emotionally from the complex reality within the atmosphere and from the basic dynamics of flying. The pilots are forced to pay most of their attention to detached knowing-that the plane will follow certain high-level orders, instead of being continuously involved with knowing-how to make the plane do what is desired (Ryle 1945, Dreyfus and Dreyfus 1988, for the difference between knowing-that and knowing-how). What appears to have happened during this disastrous event (a longer analysis follows in ch. 12.3) is that the computerized abstraction of both feedback and control-measures made the pilots unaware of certain fundamental dynamic aspects of flying, and unknowingly they stalled[10] from about 38.000ft (more than 11.000m) all the way down into the ocean without making any consistent stall-recovery attempts. The murky shadow of neat abstraction and fancy graphical representations is a potential loss of detailed insight, which ought to be a well-known fact. In this case, however, the risk for insufficient pilot insight had been addressed during design by even further abstraction and reduction of pilot influence, all in the spirit of computerized predictability being preferable to human emotions-driven irrationality. Pilots are considered more likely than a computerized flight-control system to put an aircraft into an unsafe flight mode and therefore it is considered safer if the computers fly the aircraft. Paradoxically, this enforced non-involvement implies also that pilots become even less able to recover from the unsafe situations they are accused of risk creating, resulting in such a situation becoming much worse if it would happen anyway despite all precautions, as it did for AF447. Seemingly, as the result of unfortunate weather conditions, the autopilot systems were shut down gradually in a manner that eluded the pilots to gain knowledge about what was happening, partly because it was not supposed ever to happen. The simulators where the pilots trained emergency procedures were apparently unable to enter into low-speed stall,[11] presumably because the digital flight-control system was designed to prevent this from happening thus making it unnecessary to train for. The message from the manufacturer (Airbus) was that the aircraft could not stall, a message that apparently was adopted wholeheartedly by the airline community, in this case Air France. Hence, the pilots had no training in

[10]Stall is when the airflow over a normal subsonic wing have a too large angle of attack. It happens for instance when an aircraft reduces its speed too much while trying to maintain the altitude by gradually lifting the nose (increasing the angle of attack). The result when passing over the stall edge is reduced (or more likely a complete loss of) lift and increased drag that further reduces the speed, a vicious circle that rather quickly makes the plane start falling. The only way out of this situation is to reduce the angle of attack and gain speed, either by diving or by using (a lot of) engine thrust (the latter may be possible with powerful fighter aircraft but is not really applicable to airliners), which means that stall-recovery has to be made while still having sufficient altitude.

[11]The simulators were, according to the accident report recapitulated in ch. 12.3, at least not used to train stall recovery procedures, only to demonstrate the on-set into stall.

identifying stall, or in recovering from it. On top of this, they were operating a system purposely designed to keep them from doing what actually would have been required for sorting out this particular situation and avoid the crash, which would have been to fly the plane manually out of the unsafe flight mode. All together, the position of these pilots, the position of having the responsibility to maintain safety without proper means to do so, appears as an unsatisfactory work situation, if you ask me.

The calculative approach to safety clearly distinguishable in the build-up of the AF447 disaster did evidently not result in a safe system, and because the purpose of an airliner is to transport people safely through the air one could righteously say that the technology was not sufficiently useful. It was at least not useful for recovering from stall and getting out of this kind of dangerous situations, despite the fact that stall is a fundamental aspect of aerodynamics and that being close to stall is rather frequently occurring for airplanes (aircraft are, for example, balancing on the edge of stall when landing). The AF447 disaster is, apparently, an example of a fatal accident caused by the brittleness that comes from judging technologically enforced predictability and abstracted stereotypic system operation safer than judicious system control based on experience, knowledge, and skill.

The air transport system is ultra-safe. There are, in comparison with other transport systems, extremely few accidents with airliners. Nuclear power plant accidents are, fortunately enough, also very rare. Unfortunately, the safety in ultra-safe activities has a high risk of becoming brittle, simply because the lack of incidents is a calculative result that effectively opposes the building of experience required for maintaining a generative safety.

1.4.5 The analogy between safety and usefulness

The relations between safety and resilience as well as that between effectiveness and potentiality are in a sense analogous to the relation between the contemporary model-based view of usefulness and the contributed situated view of usefulness. The notion of brittle safety should then, however, be swapped for stereotypical usefulness, and fatal accidents replaced by reduced autonomy, the latter being the ultimate undesired consequence of a calculative approach, rhetorically exaggerated in the plateau metaphor. Being blindfolded, thereby deprived from means for situation assessment and intervention, inhibits people from consciously and with personal responsibility for their own actions walk as close to the edge they consider themselves able to do without taking unnecessary risks according to their own judgments. The plateau-people are instead, by technological means in the form of blindfolds, forced into a stereotypical behavior, varying randomly around the center of the plateau. Control authority is removed from local contexts and transferred to external powers focusing solely on global aspects. The beneficial lowering of a general

probability for falling off the edge is achieved by substantially reducing individual freedom of choice. Freedom is reduced not only by a direct removal of means for situation assessment (i.e., eyesight), which are means required for deliberate interventions, freedom is also reduced indirectly because the obstructed means make it practically impossible to develop the incentive to diverge from the general. Clearly, in conjunction these reductions result in a thoroughly crippled autonomy. Without knowing about options there is no reason to diverge from the norm and the result is heteronomous behavior. Computerized systems often tend to have the same effect as a blindfold because computers do not convey naturally aspects required for alternative interpretations, which presumably is unintentional in most cases. Computers convey only what they are explicitly designed to convey, which is what is judged necessary to convey according to assumed working conditions and purposes of use, conditions and purposes that cannot be anything else but model-based predictions. Essentially, people are thereby forced to align their behavior with the stereotypical view of the world that is implemented in software because few can know enough about system workings to develop the incentive to do otherwise.

The analogy between safety and usefulness is thus that both concepts cannot be properly described only from a detached perspective. Both concepts require involved and contextual meanings for completeness. A major difference is, however, that insufficient safety is tangible and after the fact often measurable in hard counts of effects that are uncontroversial to designate as undesired (i.e., accidents), while insufficient usefulness implies stereotypical effects that, within a paradigm obsessed with predictability and that mistakes model-consistency for appropriateness, tend to be considered the desired.

The purpose of the analogy is to make the problem with model-based usefulness comprehensible and sufficiently urgent to initiate a social change. By relating usefulness to the more intuitively graspable concept of safety due to its tangible consequences, by comparing stereotypical usefulness and lack of potentiality with brittle safety and lacking resilience, my aim is to counteract the transformation of people into clueless model-followers, a situation following from computerized blindfolding. Furthermore, besides implying that we give up on our autonomy, the development seems in addition to make people refrain from taking responsibility. In some sense, that is an understandable reaction. If a person does not know how to make a difference, then there is no use trying, is it? Why should someone make an effort to do anything right, when the system is the one always doing the right thing anyway? The question is if we want to be that gullible. Does the map always show us the right way?

Model-based aspects such as stereotypical effectiveness and calculative safety are important but not sufficient, together they make up merely one dimension of usefulness. Generative aspects such as potentiality and

resilience make up another dimension, required for situated usefulness. The generative dimension facilitates contextually meaningful effects. It supports human autonomy and a natural taking of responsibility for system effects. Lack of a generative dimension is therefore a far worse situation than lack of the calculative dimension because the former manipulates our understanding of things. Clearly, lack of calculative usefulness makes a system useless in a concrete sense, while the lack of generative usefulness makes a system appear useful despite its shortcomings. For situated usefulness, a situated controllability is required because we human beings understand the world much by interacting with it. Knowing how and knowing why to control our systems in local situations is therefore essential for situated usefulness because otherwise we do not really know what is desired. Edge awareness comes from knowing how and edge awareness makes you know why. Without edge awareness, we are unable to distinguish stereotypical model-based effects from situated and contextually relevant effects. We become clueless and gullible without edge awareness and systems that makes us gullible should not be considered useful, which ought to be equally obvious as that unsafe systems are useless, hence the analogy.

2 Research approach

2.1 Philosophy

Why bother with philosophy? According to Collier (1994, p. 16, emphasis in original), “A good part of the answer to the question 'why philosophy?' is that the alternative to philosophy is not *no* philosophy, but *bad* philosophy”. Everyone without an explicitly thought through philosophy apply in practice an unconscious philosophy, of science or politics or daily life. There are always underlying assumptions about, and conceptualizations of, the world. These assumptions define what things can be in relation to each other. They define the ontology. Moreover, these assumptions define what can be known and how things can become known. They define the epistemology. Consequently, philosophical assumptions define methodologies possible to use. They define how research findings can be validated. Hence, it is necessary to begin describing the philosophy.

Within the social sciences, a war has long been raging between the paradigms of positivism and interpretivism (Mingers 2004), or between empiricists and idealists, or between realists and relativists, or between one label and the other. All while any categorization under such labels and the setting-up of a dispute between two of them probably is to make things much too simple. They are probably right in some sense all of them. One wonders therefore, is it possible to identify a core issue causing the disagreement? It appears, actually, that what matters, and what therefore is disputed, is the assumed precedence of ontology and epistemology. Does the constitution of the world – the ontology, define what we can know about it – the epistemology, or does the constitution of our knowledge – the epistemology, define what we can know about the world – the ontology?[12]

Ontology first – epistemology after: The realist builds on that the world is independent of our knowledge of it. The empiricist adds that the only thing we can know about the world is what is observable. The positivist maintains that it is possible to collect objective observations about the world. The realist-empiricist-positivist scientific goal is to generate predictive universal theories. Knowledge is created with hypothetical-deductive methods. Findings are assumed true until falsified and the theories are tested and

[12]The following two short characterizations of positivism and interpretivism are based on a clarifying summary of the paradigms provided by Wynn & Williams (2008).

verified empirically. Positivism relies heavily on quantitative research and mathematics is considered the language of science.

Epistemology first – ontology after: The relativist considers reality constructed by subjective human interpretations. The idealist adds that the only thing we can know anything about is our own thoughts. The interpretivist maintains that all knowledge and observations are subjective. The relativist-idealist-interpretivist scientific goal is to generate understanding. Knowledge is created by scientists learning to understand the meanings and actions or subjects studied. Interpretivism relies heavily on qualitative research.

This admittedly fragmentary description, a caricature of the warring paradigms, has one main purpose, to provide some kind of frame of reference for the philosophical territory within the social sciences because the underlying conceptualizations of the world, the philosophical assumptions for the present work, did not align very well with either of these two major paradigms. On the one hand, hard-earned experiences of the harsh reality made the realist stance self-evident (e.g., ch. 3.3). On the other hand, salient experiences of the indubitable subjectiveness of things enforced a relativist stance (e.g., ch. 3.4 & 3.5). Undoubtedly, this position made things slightly complicated. Fortunately, others have experienced a similar dilemma, and for some time now, *critical realism* appears to provide a suitable philosophical framework to sort out the inconsistencies (e.g., Dobson 2001, Mingers 2000, 2001, 2004, Smith 2006, Wynn and Williams 2008, Mingers *et al.* 2013). I found the scientific journey from hard systems thinking and formal mathematics, via phenomenology and soft systems thinking, described by Mingers (2004) having remarkable similarities with my own development towards critical realism. One common denominator appears to be personal experience from working with dynamic real-world issues. For Mingers these experiences came from Operations Research and management, for me from Human Factors in aviation. Such experiences appear significant for considering CR philosophical assumptions interesting.

2.1.1 Critical realism

The philosophy that has come to be called critical realism (CR) may in fact be thought of as two philosophies, or as one general philosophy with two significant applications. Primarily it is a realist ontology and general theory of science named transcendental realism by its principal founder Roy Bhaskar (2008).[13] In addition, critical realism is a theory about

[13]The referenced book by Bhaskar, *A Realist Theory of Science* (RTS), was first published in 1975 and he has written several more books on the subject. What soon will be evident, the primary CR source has for this work been, apart from RTS and a number of scientific articles (some referenced), *An introduction to Roy Bhaskar's Philosophy* by Andrew Collier (1994), which besides RTS covers: *The Possibility of Naturalism* (PN) from 1979, *Scientific Realism*

transcendental realist implications for the human sciences, an approach called critical naturalism. The name of the philosophy is thereby in some sense shorthand for these two applications in conjunction, a notion that might not be optimal but have come to be the established reference to the philosophy in question, and a notion retrospectively accepted by Bhaskar himself (Collier 1994, p. xi).

The trouble with scientific realism is twofold. Realism is "too obviously true to be worth saying … [and] anything so obvious to commonsense is probably false" making realism often be rejected on behalf of both these conceptions, dismissed as obvious and replaced by a supposedly less naive non-realist account (Collier 1994, p. 3, brackets added). Critical realism is, arguably, neither naive nor obvious (as in trivial), it is about acknowledging both the fact that realism must obviously be true and the fact that our knowledge about the real world inescapably is subjective and relative. The term realism is there to express clearly the assumption of realist ontology as well as the stance that ontology comes before epistemology. For a critical realist, the constitution of the world determines what we can know about it, not the other way around. This conclusion is drawn, if for no other reason, simply from the fact that the world predates human beings that presuppose the presence of the world, which means that the world exists independent of our conceptions of it. The term critical should be interpreted in its philosophical sense with a positive connotation, contrasting with naive or dogmatic, thus appeal to the relativist stance. Overall, CR is more about providing philosophical means for joining the warring parties (e.g., positivism and interpretivism) than about carving out yet another stronghold from where the other positions would be fought. However, there is one critical aspect that CR, here taken as represented by Bhaskar (2008), that Collier (1994) and Archer et al. (2007) goes through some trouble to refute, which is the Humean notion of causality as a constant conjunction of events, a notion used by empiricists to predict outcomes. Bhaskar calls this view Actualism (e.g., 2008, chap. 2), which reduces the real world to be merely that of the actual. While the studying of arguments for these rather elaborate refutations of strong empiricism and positivism as well as of super-idealism and relativism definitely is an interesting way to learn about CR (e.g., Collier 1994, chap. 3, The Impossibility of Empiricism and Idealism), it is beyond the scope of this research to recapitulate this discourse at length. For the present purpose, an overview of the central aspects of CR will suffice.

The core aspect of CR is that it recognizes a depth both in reality itself and in our knowledge about it, there is an ontological as well as an epistemological depth. Let us begin with epistemology and the depth

and Human Emancipation (SRHE) from 1986, *Reclaiming Reality* (RR) from 1989, and *Philosophy and the Idea of Freedom* (PIF) from 1991. Unfortunately, I have not found the time (yet) to read all these books myself. However, as a complementary source, *Critical Realism – Essential Readings* (Archer *et al.* 2007) has also been (briefly) consulted.

dimension of knowledge. As a realist philosophy, CR concludes that the object of science is real, independent of the science itself. Science is about something and science can be wrong about its object. Hence, there is a difference between the object of science and knowledge about this object. For example, science can know about the speed of sound, and this knowledge might be sufficiently accurate. The important thing, however, is that real sound waves travel at some speed regardless the accuracy of the scientific knowledge about it. CR defines the object of science the intransitive object (i.e., the real object) and the current knowledge about this object the transitive object (i.e., the relative object). The current knowledge about the object of science (the transitive object) is the raw material of science, which by scientific work is constantly transformed into a deeper understanding of the intransitive object. The term transcendental, often connected subconsciously with spiritualism and other supernatural aspects (of which transcendental realism has nothing to do), is perhaps slightly more intelligible when these two sides of knowledge is understood. Critical realist science aims to transcend the limits of our knowledge, as reality (the intransitive domain) always is more than we know (the transitive domain). This usage of the word transcendental is similar to how Kant used it in the philosophy of transcendental idealism, but quite in contrast with its content because in the Kantian philosophy it is ideas that transcend to reality.

For a critical realist the scientific question is transcendental in the opposite direction compared to Kant, as in: what must the world (the reality, the intransitive object) look like to make our understanding of it (the idea, the transitive object) possible? This kind of question is required in order to transcend the gap between the intransitive domain of reality and the transitive domain of knowledge. There is, however, a complicating fact, a paradox, for science in general and for critical realist science as well, which is that science can never be assumed correct. While any scientific effort necessarily must aim for a perfect understanding (a half-hearted attempt would simply not make sense), the goal of perfect knowledge is forever unreachable. This fact can be explained by the distinction between the intransitive and transitive dimensions because it depicts why science is fallible. Science is fallible because the transitive object will never be the same as the intransitive object. Even if the transitive object would happen to be perfect, there is no way to know that this is true. The work of science is therefore eternal.

In a single phrase, the ontological assumption of CR may be described as stratified depth realism. The first dimension of this two-dimensional depth is illustrated by describing three domains: the domain of the real (D_R), the domain of the actual (D_A), and the domain of the empirical (D_E). The second dimension concerns stratification and emergence, aspects discussed in the following two sub-chapters (ch. 2.1.2 & 2.1.3). Ontology comes before epistemology, and the fallibility of science together with the subjectiveness

of observations make possible knowledge (transitive objects) from observations (empirical knowledge) always be a subset of what may be observable (actually occurring), which in turn always is a subset of what can occur and is possible to know about (intransitive objects). That is, $D_R \geq D_A \geq D_E$. The domain of the empirical consists only of experiences (observations). These experiences occur, and they are generated by events that occur, in the domain of the actual. Experiences and events are also real, thus belonging to the domain of the real as well, in which there are also mechanisms that produce events that may be experienced. This is depth realism, summarized in Table 2.1 below (adopted from, Bhaskar 2008, p. 56).

	Domain of Real, D_R	***Domain of Actual, D_A***	***Domain of Empirical, D_E***
Mechanisms	√		
Events	√	√	
Experiences	√	√	√

Table 2.1: The domains of critical realism

The notion of a deeper reality implies that the empirical reality is not the entire truth. There might actually be unknown underlying real events that create (cause) empirical experiences interactively. Just as intransitive objects exist independent of the transitive objects of knowledge about them, real events can occur without being experienced, and real events may exist that cannot be inferred from empirical observations only. Consequently, the domain of the empirical is dependent on but not the same as the domain of the actual, implying that empirical experiences can be explained by actual events while the reverse does not hold. It is impossible to experience real events not existing within the domain of the actual. Hallucinations and other illusions purporting to exist in the domain of the actual may be real experiences, but they are not experiences of real events independent of the individual experiencing them. Empirical experiences do not cause actual events to happen, events possible to be experienced by other subjects. Analogously, experiences can neither explain events, unless the transcendental construction is used, thereby subject to the fallibility connected with it. Therefore, the domain of the empirical is less than, and only as a special case equal to, the domain of the actual.

This kind of relation, as that between the empirical domain and the actual, is then repeated for the domain of the actual in relation to the domain of the real. While actual events indubitably are real, and inferred mechanisms explaining these events are plausible, there might be more to the world. There may be additional underlying real mechanisms producing events, mechanisms that cannot be inferred from empirically observable events only, neither by adding events conceptualized by transcendental arguments (i.e., from the domain of the actual to the domain of the empirical, e.g., by asking

what must actually have happened for the empirical experience to be true?). For a more complete or at least less limited knowledge about mechanisms (although still a transitive knowledge), the transcendental reasoning must be applied here as well. Furthermore, CR acknowledges that there may be a multiplicity of mechanisms at work simultaneously, resulting in the fact that events can be explained in several perhaps contradictory ways. In particular, there may be mechanisms existing for real, but unexercised or canceling the effects of each other, thereby not creating actual events. Thus, the domain of the actual is less than and only as a special case equal to the domain of the real. Therefore, one should be careful about stating that something is unreal (i.e., does not exist) simply because it is not concrete enough to cause real events readily observable by empirical means. Critical realism is, in an effort to summarize the above, sometimes characterized as ontologically bold, but epistemologically cautious (Wynn and Williams 2008).

2.1.2 Openness, explanations, and the epistemic fallacy

Theoretical systems are constructions for understanding. They are transitive objects describing intransitive objects. Systems can either be open as in affected by external factors or closed as in not affected by external input. Within a closed system, there can be no causes and effects, only conjunctions of phenomena or event regularities. Hume pointed out that such event regularities are all that ever can be observed, a conclusion based on the assumption that the world is a closed system. According to CR, however, in practice, there are no closed systems and therefore the observation of events and induction of laws producing these events are not enough.[14] Steele (2005, pp. 138, 144, respectively, brackets added) criticizes CR for being "erroneously empathetic in rejecting the scientific relevance of event regularities", but acknowledges at the same time that a "closed system exists only as a theoretical device … [because there] is no practical illustration of one" and appears therefore to have missed the essential point of the argument about openness and depth reality. It is not that the orthodox deductive-nomological (ODN) tradition and observation of event regularities is totally rejected, on the contrary, CR acknowledges its relevance and aims to join the strengths of ODN with qualities of more relativist methods. The problem is, according to CR, the conclusion that knowledge gained from observing event regularities defines reality. That conclusion is a dead end for

[14]The assertion about inescapable openness may obviously be disputed, but its defending is deferred (handed over) to the discourse within CR literature refuting strong empiricism and positivism. For, arguably, positivism relies fundamentally on the assumption that reality in fact is a closed system (or can be generalized as one), and any openness is therefore merely a consequence of not having complete knowledge about the system. Whether reality in fact is closed does not matter for the present purpose. It suffices to realize that relevant natural systems in practice are open and that the critical realist ontology therefore provides practical guidance avoiding certain crucial errors that tend to occur without it.

science and something Bhaskar (2008, p. 36) has named the epistemic fallacy, which "consists in the view that statements about being can be reduced to or analyzed in terms of statements about knowledge; i.e., that ontological questions can always be translated into epistemological terms". "For the transcendental realist it is not a necessary condition for the existence of the world that science occurs. But it is a necessary condition for the occurrence of science that the world exists and is of a certain type" (Bhaskar 2008, p. 38). Moreover, this world happens to be of a type where the domain of the real is greater than the domain of the actual that is greater than the domain of the empirical. The epistemic fallacy, however, has arguably been around for a while:

> Since Descartes, it has been customary first to ask how we can know, and only afterwards what it is that we can know. But this Cartesian ordering has been a contributing factor to the prevalence of the epistemic fallacy: it is easy to let the question how we can know determine our conception of what there is. And if in a certain respect the epistemic question does seem prior, in another it is secondary to the ontological one: knowledge exists as an aspect of our being in the world, and before we can know how we know, we need to have some idea how we interact with that world in such a way as to acquire knowledge of it (Collier 1994, p. 137).

The inescapable openness of systems can therefore be seen as a direct consequence of the limitations of knowledge. System descriptions are models of the world that are defined by the analyst and as such they are epistemological not ontological entities (Vicente 1999, p. 77), which means that when discrepancies between model and reality occur, it is generally impossible to distinguish between external (truly open) factors and internal misrepresentations (model errors).

Experiments, an essential scientific activity, fundamental to the ODN tradition, are deliberate efforts to create artificially closed systems. It is at the core of experimentation to address the problems with extraneous variables (open factors) and threats to internal validity (model errors). The purpose of system closure is to isolate the mechanism studied, preferably only one. Knowledge (i.e., the theory) is then validated by confirming predictions or hypotheses of system effects. Experiments are explicit efforts to achieve $D_R = D_A = D_E$ such that inferences, or generalizations, later can be made about the domain of the real from the observations made in the domain of the empirical. The present stance is that there is nothing wrong with this approach. It is probably a necessary approach for achieving the rigor required for detailed and explicit knowledge about the mechanisms studied. However, when used to argue that everything can and must be explained this way (positivism), implying that what cannot be explained this way does not exist for real (strong empirical realism), the epistemic fallacy is in effect, a situation associated with certain undesired consequences. These empiricist-

positivist arguments are based on the erroneous assumption that the artificially produced special case, the case when the domains are equal, spontaneously occur in reality and therefore can function as proof of general validity for empirical inferences.

One consequence of considering explanations by theoretically closed systems as not only valid but also the most valuable because they stand as the objective truth, is that it enforces, or encourages, a detached perspective. The closed system must be, or is at least preferably, observed from the outside. Otherwise, the system includes the observer, which tends to make it obvious for the observing analyst that the observed system must also incorporate personal matters and social aspects thus quickly grow to become incomprehensibly complex and practically unanalyzable with satisfactory rigor. Because of the detached outside perspective, the analyst tend to lose touch with situated (involved) matters (more about this below in ch. 2.1.4 & 2.1.5). Presumably, a loss of the involved perspective is not a problem when studying matters covered by physics and other natural sciences, as it should be quite safe to assume that physical particles do not care about whether there are alternative viewpoints. However, when studying psychological and social aspects, the lost perspective might include what really matters, especially for whom it concerns, meaning the participating human elements of the analyzed system.

If constant conjunction of events identifying natural necessities (i.e., Humean causes) are insufficient for explanations of objects from the domain of the real, what are CR explanations then made up of? There are four concepts that make up the transcendental realist theory of natural necessity (Collier 1994, p. 61): structures, powers, generative mechanisms, and tendencies. To elaborate on these it is first necessary to look into the other dimension of the CR depth reality, the stratification, and scrutinize the concept of emergence.

2.1.3 Stratification and emergence

Besides the deep reality in which a multiplicity of mechanisms open up for new events and explanations relevant to a certain phenomenon forever to be uncovered, critical realists acknowledge that reality is also stratified. Stratification implies that mechanisms belong to certain layers or strata in which they can be explained and outside which they make little sense. Collier (1994, p. 107) explains the stratification of nature by comparing it with the stratification of science. In some sense, everything can be studied by physics and every material substance can be studied by chemistry. Only some of these material things can be studied by biology, and yet a subset of these biological things can be studied by psychology. All animals are made up of chemical substances but not all chemical substances are present in animals. This relation implies that strata are ordered. Animals cannot ignore

the laws of chemistry and physics and thus they belong to higher strata depending on the lower ones.

The stratification of nature can also be explained as a stratification of mechanisms or laws. Biological mechanisms cannot exist without chemical ones, while the reverse is not true. There are chemical mechanisms that have nothing to do with biological laws, for example, chemical laws for chemical substances not present in biological objects. There are also biological laws not applicable to chemical substances (e.g., the law of natural selection), which then is the rationale for biology science studying such biological mechanisms. The ordering of strata and the direction of dependencies are important aspects. Bhaskar (2008, p. 113) makes a distinction between rooted in and emergent from. In philosophy of science, the concept emergence explains things that are "neither predictable from, deducible from, nor reducible to the parts alone" (Goldstein 1999, p. 57). Biology is rooted in chemistry because it is governed by the laws of chemistry, while having emergent biological mechanisms irreducible to the mechanisms of chemistry. Emergence is what makes it inappropriate to consider biology redundant to chemistry and chemistry redundant to physics.

Emergence is, however, a much-disputed concept, essentially attacked from two directions. Emergence is rejected by pluralists that argue for no connection at all between higher strata and lower, and it is rejected by reductionists holding that there is only one real stratum (Collier 1994, p. 111). Once again, for the present purpose the true answer to this highly delicate question does not matter, but arguably, neither of the attackers bring arguments that are of any particular use. For the pluralist approach that argue for no connection there is nothing relevant to learn for higher strata from lower, implying for example that the mind of human beings cannot be studied by neurobiology, while the reductionist approach is apparently insufficient for explaining certain emergent aspects of higher strata. CR ontology, on the other hand, provides a fruitful arena to explore.

By acknowledging the stratified nature of reality, the rootedness of higher mechanisms in mechanisms from lower strata, the irreducibility of emergent mechanisms to those of lower strata, as well as the deep reality with the three domains, two kinds of explanations become relevant. One or more mechanisms of lower strata explaining a higher mechanism becomes a vertical explanation, whereas a mechanism plus a stimulus (a trigger) that explains an event is a horizontal explanation. It is important to keep in mind the irreducibility, for otherwise vertical explanations could be seen as explaining away the higher, or in opposite terms, the higher mechanisms could then be reduced to and considered fully explained by the lower. Bhaskar (2008, pp. 114–116) expounds on the impossibility of such reduction by stating that no science can be reduced to that of the lower, often regarded as purer, without a prior body of knowledge about what is to be explained, a body of knowledge produced by the science to be reduced.

Therefore, as "means for discovery, i.e., of achieving such a body of knowledge, reductionism must fail. For it presupposes precisely what is to be discovered" (ibid, p. 116). Whereas CR maintains a continuous evolution of changing strata, Emmeche et al. (1997) suggests four primary levels of emergence, in which there may be sub-levels. These primary levels are the physical, the biological, the psychological, and the sociological.

The four words for explanation of natural necessity can now be described as belonging to different strata relatively, where structures are at the more basic stratum, explaining vertically the powers of mechanisms at higher strata and so forth. The reoccurring and perhaps most concrete example of a relation between structures and powers given by Bhaskar (2008) and Collier (1994) is that of chemical reactions. The structure of the molecules of substances is what gives them the power to react. However, for a reaction to occur in practice, these powers must sometimes be triggered, usually by adding a necessary condition (e.g., heat). However, these relations become more complicated when reaching higher strata, especially when entering the domain of the social sciences, all while CR argues that there are social structures as well as social powers, generative mechanisms, and tendencies. The complexity of social strata may be illustrated by considering the relation between psychology and sociology, both rooted in (but not reducible to) biology and physiology, mutually dependent yet depicting distinct mechanisms. Social mechanisms cannot exist without psychological mechanisms and, arguably, vice versa, many psychological mechanisms cannot exist without social relations. All the four words are in some sense causes, as in horizontal explanations, or mechanisms producing events and experiences.

To illustrate all four kinds of mechanisms, and especially the last two not yet discussed, the social phenomenon of a riot might help. Social structures can be seen as a framework shaping ideas of and relations between individuals. Desires within individuals, arguably quite significantly influenced by social relations, are powers that might be triggered to foster actions and activities. Within these social structures, activities and desires can form generative mechanisms that might whip up the rage of a certain group of people about something they care for, and sometimes such a development evolves into a riot. Finally, there is a tendency that riots lead to damaged property. The four explaining words designate mechanisms that are part of an overall explanation for why some property became damaged, but a riot is not clockwork machinery. It is not possible to pinpoint a social structure plus a desire that will generate a riot that will damage certain property, because, the social system is not deterministic (i.e., it is not closed). Such pinpointing, however, is often the result of a detached perspective with hindsight where systems tend to be assessed as they were closed, in the form of predictions with statistical significance. Yet, the structures, the desires, the generative mechanisms, and the tendencies, are

real phenomena important to understand, phenomena that are irreducible to, say, neurobiology or sub-atomic physics. They are emergent phenomena that lose most of their relevance when studied and described with a detached perspective that, if described such, then become theoretical descriptions with little practical value.

2.1.4 Knowledge and practical wisdom

Philosophy is the foundation on which knowledge resides because philosophical assumptions determine the character of possible knowledge. Despite that knowledge about knowledge may seem somewhat academic or philosophical, it appears reasonable to state that knowing about the nature of knowledge can facilitate a more appropriate application of possessed knowledge thereby improving its potential usefulness. It is reasonable in the same way as that knowing about the nature of wood can facilitate a more appropriate utilization of possessed timber thereby increasing the probability for wooden constructions to become successful. Appropriate utilization of timber includes, naturally, also a certain amount of concrete skills and craftsmanship, which arguably is true also for utilization of knowledge. Theoretical knowledge is of little use without practical skills to use it.

Practical wisdom is discussed here because it is a philosophically interesting concept that relates to scientific knowledge in a manner similar to how situated usefulness relates to technical conditions. A too strong focus on detached formal descriptions such as explicitly depicted knowledge implies a reduction of relevance for involved emergent phenomena and tacit knowledge in the same way as a too strong focus on specific performance measures frames expectations of usefulness to aspects affected by the measured properties. The following discussion serves also the purpose of being an elaborate introduction to the involved perspective.

The practical dimension of knowledge is of prime interest for the present research simply because knowledge is a prime facilitator of usefulness, and practical skills to apply this knowledge is argued to be what allows for usefulness to become situated. Practically knowledgeable people are often attributed as wise, indicating the possession of reassuring quantities of useful knowledge as well as skills to use it. So, what is practical wisdom then?

Aristotle was one of the great ancient Greek thinkers and he was a member of Plato's academy in Athens. While at large following the tradition by asking the same sort of questions as his famous tutors, Aristotle contrasts with Socrates and Plato in some aspects. One of these aspects is his emphasis on virtuous activities, not merely on possessing virtues (Aristotle 2000, p. ix). The contrasting lies in that Aristotle was concerned with the situated nature of human virtues, not merely with their idealistic and theoretical representations. The contemporary view of usefulness, the calculative technical rationality kind of usefulness, based on an exaggerated

belief in models and attainability of predetermined objectively correct courses of action, can be thought of as a tendency to align with the Platonic world of pure ideas (i.e., to prefer the map before reality). For the present purpose, the Aristotelian contrasting with the idealistic world of Plato is therefore of the greatest interest.

The ancient Greek thinkers appear to have been quite preoccupied with the soul, a concept that modern science in principle has abandoned. Aristotle, on the other hand, made great efforts to categorize the different virtues of the soul. These categories, although initially and specifically intended for the now largely obsolete soul appear, however, still relevant for describing different kinds of human qualities. Perhaps is the Aristotelian explication of human virtues more relevant than ever? It seems that along with the scientific abolition of the soul-concept, the rich and situated meaning of concepts describing certain human virtues have become reduced to denote merely formally describable idealistic and rigorously modeled properties. Concepts such as consciousness and intelligence, trust and responsibility, honor and autonomy, etc., are today by some people considered made up of properties providable by technological means. Arguably, however, technology and technological properties can merely provide idealistic and stereotypical representations of these virtues and qualities that formerly were related to the soul (elaborated on in ch. 15.1).

If one reflects on what a soul-concept might be in modern terms, it seems actually fit a need. While it is perfectly appropriate for some purposes to consider the human body a decomposable mechanical machine, for example when splinting a broken leg, issues such as psychosomatic effects and many psychological phenomena are quite difficult to explain with similar mechanistic notions. In fact, many concepts for characteristic human qualities and abilities, especially psychological qualities such as consciousness and intelligence, are rather difficult to describe in terms of concrete facts and explicitly describable properties. Human qualities seem to require a more holistic viewpoint. By using the critical realist notion of strata and considering the human being made up of layers of systems with emergent properties, a modernized version of the soul-concept could then be that 'the soul' encapsulates those features that are emergent in a bio-chemic-mechanic body-system as a whole. Chemistry, biology, physiology, certain strands of medicine – areas adhering to the powerful systemic paradigm of the natural sciences (technical rationality) – are perhaps sometimes working at a too specific level of abstraction where certain features of the human nature cannot be explained without missing the essential point. In light of the four primary levels of emergence suggested by Emmeche et al. (1997), the emergent soul becomes a great metaphor for essential holistic human qualities. Properties emergent in the physical, the biological, and the psychological primary levels, but not explicitly in the fourth sociological level, may then be considered to make up a modern human soul, which,

because of the nature of emergence is irreducible to biological, physical and technological properties. Aristotle's categorization of the virtues of the human soul can then be used as a presumably quite well thought-through outline of such emergent human qualities.

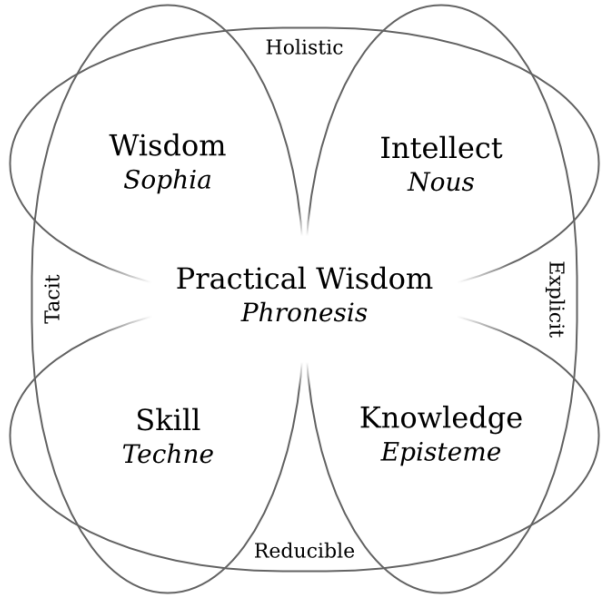


Figure 2.1: Aristotelian-inspired categorization of human virtues

The categorization of human qualities in Figure 2.1 is inspired by Aristotle's categorization of those virtues of the soul that he considered governed by reason, outlined in Book VI of the Nicomachean Ethics (Aristotle 2000, pp. 103–118). The fact that Aristotle was considering only virtues of the soul and that the present categorization is considering virtues of the human being as a whole make, however, some essential label-concepts have slightly different meanings compared to the ancient interpretations. For example, when describing skill (techne) the original concept was merely about skillful producing of reason, although, that kind of reason was considered a prerequisite for deliberate and skillful production of real-world things. That description should be contrasted with the present interpretation of the concept of skill that also includes real-world production skills. Thus, skill is here considered both a virtue of reason and of bodily performance. The same goes for action and activity, for which Aristotle in his categorization was considering the taking of active stance in matters and the making of decisions. Actions and activities are here considered essentially physical and occurring in the real world, thereby necessarily performed in a manner that includes the body, but governed by holistic psychological aspects (e.g.,

Figure 6.1). Furthermore, the grouping of concepts is made according to aspects considered relevant for this work, thereby not necessarily an orthodox interpretation of the Aristotelian philosophy, which on the other hand never was the aim.

The first category is scientific knowledge (episteme) that is based on what is already known, thereby teachable and learnable. Scientific knowledge is here described as reducible as in systemically decomposable into bits and pieces, and explicit as in describable and representable. The decomposability makes scientific knowledge teachable because that is what enables complex knowledge to be conveyed in graspable portions. The explicitness makes it intellectually learnable because that is what enables the conveyed portions to be joined in a structured manner to construct new knowledge. The explicitness is in addition what makes the quality of the transfer testable by use of definable measures (e.g., scores on exams).

The second category is skill (techne), which is about the deliberate production of things. Skills can also be considered reducible because they are often systemically decomposable into sub-skills, thereby describable in terms of their separate goals and functions. However, most skills, and especially the embodied portion of concrete skills, are mostly tacit (Polanyi 1966) as in essentially non-describable. It is for example possible to describe the abstract principles of riding a bicycle, a description then falling into the category of scientific knowledge, but the core bicycle-riding skills remain tacit. It appears as if certain skills are adopted by the body as a whole, not filtered through deliberate reasoning, thereby inaccessible to explicit descriptions. In addition, tacit skills tend to leave no other marks than a firm gut feeling, in this case consisting in a certain amount of bicycle-riding confidence. Tacit knowledge such as skills cannot be readily taught and thus not learned in a traditional academic (or scientific) sense, but must instead be learned by trial and error for the possibility to gain personal experience. However, such learning is probably a process benefiting from teaching, analytic (scientific) reflection, and practical guidance.

The central category is practical wisdom (phronesis) that is neither knowledge nor skill, yet it is both. Aristotle states, “We may grasp what practical wisdom is by considering the sort of people we describe as practically wise” (2000, p. 107). In the present context, this means people such as successful and proven trustworthy commanders, physicians, pilots, captains, and the like. People that always seem to get things right in their line of work, almost regardless the challenges they face, in their particular area of expertise of course. They do not only know a lot about what they are doing, they also know how to do what they do, and perhaps most important, they seem to know intuitively when to disregard established procedures, go against their own theoretical knowledge, and try something new, which they in addition tend to succeed with remarkably often. Practically wise people seem to possess the ability to judge wisely among ambiguous or even

opposing facts of knowledge, select wisely a suitable course of action, and have the required skills to go through with it. From a detached perspective the success of such intuitive know-how may often seem like pure luck, but arguably, the practical wisdom this kind of people have gained does contribute significantly to the outcome.[15] An alternative word for practical wisdom could be expertise. For sea captains and pilots, the notions of seamanship and airmanship appear to align rather well with Aristotelian phronesis. Moreover, the title concept of edge awareness is considered a core aspect of practical wisdom.

The last two categories, intellect (nous) and wisdom (sophia) are actually not that important for the present work, but they are included in the model for completeness. They are described as depending on practical wisdom that, however, is relevant in this context, and perhaps are intellect and wisdom the most emergent of all emergent human features thus the most difficult ones to describe in terms of concrete facts and systemic decompositions. Intellect and wisdom are clearly holistic human features, where intellect can be considered to lean towards the explicit and scientific side while wisdom aligns with the tacit and embodied side.

2.1.5 Two perspectives, detached and involved

From the perspective of the present research, reductionism and empirical realism, especially when applied to psychological and sociological aspects, become idealistic kinds of realism. They are idealistic as in maintaining the view that certain things can be fully understood in terms of their parts, with clear-cut and formal descriptions of their relations. In essence, reductionism is the view that the world is adequately represented by the map. A Platonic approach to reality that ignores or dismisses the existence of emergent properties and Aristotelian situated aspects. Supposedly, the contemporary paradigm of technical rationality and idealistic objectivity (e.g., positivism) has earned its strong position because of its evident success within the natural sciences, perhaps particularly in physics where artificial closure of systems in fact is attainable. Striving for objectivity becomes, consequently, more or less implicitly a reduction of the assumed reality to concrete aspects observed from a distance, a reduction that occurs at the expense of knowledge about emergent aspects observable only within the actual system of concern.

It all comes down to a matter of viewpoint. Objectivity implies looking at something from the outside because otherwise judgments would not be

[15]The Swedish slalom legend Ingemar Stenmark once phrased this relation between practical experience and luck elegantly. After what appeared as a miraculous recovery during a race, a TV-reporter commented on his luck. Allegedly, Stenmark replied (in his characteristic slow northern Swedish dialect and deemphasizing manner) something like: – I know nothing about luck, only that the more I practice the luckier I get!

objective. If a matter is looked on from the inside it becomes personal, and subjectivity is then righteously assumed to prevail thus by definition making the objectivity of assessments fail. Objectivity is, by all means, a grand ambition, for suitable matters. The trouble with objectivity occurs when it is enforced on personal or personally related matters, like human qualities such as intelligence and expertise or values. When, for example, intelligence is analyzed from a distance, it becomes easily reduced to mean what is applicable also to other things viewed from the outside, for instance technological systems such as computers. Arguably, this is a rather limited kind of intelligence focusing on rationally defined problem situations. The fundamental difference in character between artificial intelligence and human intelligence, as well as between technological capabilities and human expertise, has been studied in depth by Hubert and Stuart Dreyfus (1972, 1986, 1992, 1980, 1988). They often refer to what they call the detached and the involved perspective, in order to explain these crucial differences.

The plateau-society metaphor described in chapter 1.4.3 is supposed to be an illustrative caricature example of the inappropriateness of focusing only on the detached perspective, while disregarding the involved perspective. Most notably, the meaning of concepts and values are, arguably, intelligible only from within the considered system, (i.e., by having an involved perspective). For science, this issue shows as the problem of balancing between the specific that has no meaning and the general that has no content (Boulding 1956), or as the dilemma of rigor vs. relevance (Applegate 1999, Benbasat and Zmud 1999, Davenport and Markus 1999). Just as for "The Three Absolute Limits to Knowledge … Heisenberg's Uncertainty, Bohr's Complementarity, and Gödel's Undecidability" (Yates 1978, p. R202), the rigor vs. relevance dilemma makes one desirable aspect counteract another desirable aspect. The objectivity required for rigor implies (implicitly) a detached perspective that counteracts the involved subjectiveness required for relevance.

The contemporary view of usefulness and related issues, in this research critiqued and redefined, was quite quickly associated with a tendency to favor the detached perspective. At the same time, focusing only on the involved perspective would clearly lead to other undesired issues, such as failing to take aspects into account that in fact are appropriately viewed as objectively correct, such as the Newtonian laws for everyday physical relations. Consequently, much of the present work has been focused on describing how these two perspectives are linked together in practice, which also is interpreted as an (implicit) aim of critical realism as a philosophy of science. CR proposes a framework (the stratified depth ontology of transcendental realism) for describing the detached perspective such that the impact of the involved perspective becomes describable (by critical naturalism). Nevertheless, the notions of detached and involved perspectives

have had great impact on the present research and will reoccur frequently throughout this thesis.

2.2 Methodology

2.2.1 Critical realist methodology

Critical realism has been categorized by some as a philosophy in search for a methodology (e.g., Yeung 1997) and by others as a philosophy suggesting a multiple methodology approach (Mingers 2001, 2003, Zachariadis *et al.* 2013). Collier (1994, pp. 122–123, 160–167) recapitulates and explains general methodological models called RRRE and DREI (explained shortly) outlined in Bhaskar's books. These models are general methods for explanations in open systems and within the social sciences. Some concepts behind these acronyms apply to the present research.

The epistemological situation with critical realist ontological assumptions makes retroduction the fundamental methodological concept, which should be distinguished from retrodiction (retro- + predict) meaning to use present information to infer or explain something in the past.[16] Retroduction is a concept somewhat similar to that of abduction introduced by Charles S. Peirce, a concept with vastly different interpretations reaching from arbitrary guessing to "a mode of reasoning that *justifies* beliefs" (McKaughan 2008, p. 446, emphasis in original). Retroduction involves not only the production of plausible hypotheses but also reasons why the hypotheses are plausible, allegedly much like how detectives determine what to investigate, presumably a practice thoroughly dependent on experience and practical wisdom (i.e., Aristotelian phronesis). Bhaskar states that "transcendental arguments are a species of retroductive arguments, i.e. arguments 'from a description of some phenomenon to a description of something which produces it or is a condition for it'" (Collier 1994, p. 22). However, in order to produce descriptions of plausible mechanisms some effort must presumably first be put on producing suitable frameworks for describing appropriately the phenomena. It may be required to develop novel or alternative descriptions such as models and conceptual frameworks of a reality in which the explained phenomena are possible. Such efforts are what the two first R's in RRRE and the D in DREI depicts, and for the present research, they correspond at large to phase one.

The first R stands for resolution of a complex phenomenon into graspable components. The second R stands for redescription of component causes, which are descriptions thereby presupposing established theories. Redescriptions are therefore not always appropriate for describing deeper or

[16]Explanation of retrodiction from http://www.merriam-webster.com/dictionary/retrodict

underlying explaining mechanisms, making it necessary to produce a completely new description (the D in DREI) of a conceptualized mechanism or tendency. Then follows for closed systems with well-known mechanisms, retrodiction (the third R in RRRE) to predicting theories, but for open systems with competing explanations retroduction (the R in DREI) to plausible explanations of the phenomenon is more appropriate. The last step (the E in RRRE or EI in DREI) is elaboration and elimination of alternative explanations, if possible by empirically controlled identification of the described mechanisms at work, where identification, if possible, may be connected with experimental closure for more rigorous results.

Wynn and Williams (2008, 2012) have reviewed interpretations of these general methodological models and developed five principles for conducting critical realist case study research (in information systems). These principles assume CR philosophical assumptions such as a stratified ontology and independent reality, mediated knowledge, an open-systems perspective, unobservability of mechanisms, emergence, and a focus on (multiple) explanations via (multiple, unobservable) mechanisms. The principles are:

- **Explication of events**: Identify and abstract the events (phenomena) studied, usually from experiences, as a foundation for understanding what really happened.
- **Explication of structure and context**: Identify components of social and physical structure, contextual environment, and relations (critically redescribed into a theoretical perspective).
- **Retroduction**: Identify and elaborate on powers and tendencies of structure that may have interacted to generate explicated events (phenomena). Identify plausible candidate causal mechanisms with logical and analytical support for explanations
- **Empirical corroboration**: Ensure that proposed mechanisms have causal power and that they have better explanatory power than alternatives. Based on case data, validate analytically proposed mechanism and assess explanatory power.
- **Triangulation & Multimethods**: Employ multiple approaches to support causal analysis based on a variety of data types and sources, methods, and theories.

2.2.2 On problematization and theorizing

The epistemological gap between the intransitive real world phenomenon and transitive knowledge about it makes scientific knowledge always be theoretical in some sense. Theorizing must therefore be a core scientific activity and theory the essential result, must it not? The crucial questions become then: what is theory, what is it not, and is it at all possible to separate the product – the resulting theory object – from the process consisting in the

way the theory comes to be and is used (e.g., Sutton and Staw 1995, Weick 1995)? In the present context, the last question becomes whether it makes sense to view theory strictly from a detached perspective (i.e., focusing on the product and on its objective and generalized, e.g., calculative, validity) or if it would be more appropriate to include the involved perspective and consider situated and generative aspects of theorizing and theory application as well? Presumably, both aspects are necessary, not least because empirical findings, methodological approaches, and theoretical advances, tend to depend mutually on each other. For additional guidance in these matters, it seems therefore relevant to review what is considered a theoretical contribution. "That is, what signifies a significant theoretical (as opposed to an empirical or a methodological) advancement in our understanding of a phenomenon?" (Corley and Gioia 2011, p. 12).

Good theory is plausible, interesting, obvious in novel ways, a source of unexpected connections, high in narrative rationality, aesthetically pleasing, and correspondent with presumed realities (Weick 1989). However, to know (or anticipate) what will be interesting and unexpected yet plausible and correspondent with reality is not a trivial problem, it is the problem of the problem (Weber 2003). Hence, good theory requires primarily a relevant problem for which it is theory about. While building a theory about theory building, Corley and Gioia (2011) synthesized the current view on what constitutes a theoretical contribution and found that such contributions are often judged by the two criteria of originality and utility, and added relevance to practice as a prominent dimension of theoretical contribution. Weber (2012) lists novelty, importance, parsimony and level, as criteria for evaluating the quality of a theory as a whole. Originality (or novelty) may take the form of incremental insight within normal science according to the prevailing scientific paradigm, or the form of revelatory insight where the theoretical contribution "reveals what we otherwise had not seen, known, or conceived" (Corley and Gioia 2011, p. 17). Parsimony and level, aesthetics and catchy phrasings, are also important qualities for good theory, facilitating accessibility, comprehensibility and applicability, but what about relevance, importance and utility? Theories should be important and utilizable for what, for whom, how, why, when, and where? These standard questions are "the Building Blocks of Theory Development" (Whetten 1989, p. 490), although they may imply walking down two different alleys.

With a general, and perhaps overly categorical, distinction between academia assumed focusing on science and the rest of the world assumed focusing on practice, theories will be about different things depending on affiliation. The difference is because reflection about issues as well as problematization and theorizing tend to regard pressing matters present in the everyday life of thinking individuals, thereby centering on personal interests and following procedures common to the community, which seems quite natural after all. It is therefore also quite understandable if scientists

primarily are concerned with scientific relevance and practitioners with relevance for practice. Hence, the categorical distinction implies two different kinds of utility for theories, scientific utility and practical utility (Corley and Gioia 2011, pp. 17–18). It is consequently neither surprising if scientists tend to focus on scientific utility, on improvements in rigor and testability of theories, and if they tend to look for new areas to explore by searching for gaps in present scientific knowledge. However, gap spotting and a too strong focus on the scientific aspect of utility according to norms within a particular scientific community tend to reinforce the prevailing paradigm. On the other hand, if practitioners theorize, who are non-scientist individuals thereby presumed unfamiliar with scientific procedures and scrutiny, if they theorize, then it appears likely that resulting theories will lack necessary scientific qualities that therefore are inappropriate as theory in a scientific sense. Yet, no matter the scientific qualities, theories without practical relevance are virtually useless because, frankly speaking, academia and science is not an end in itself. The goal of science is, at least as far as I am concerned, to understand the world and through enhanced understanding influence practice, an understanding that, arguably, must take different forms depending on what it is about.

2.2.3 The nature of theory

Gregor (2006) defines five types of theories: I – for analysis, II – for explanation, III – for prediction, IV – for explanation and prediction, and V – for design and action. Weber (2012) argues that type I is better viewed as typologies, type V as models, and types II and III as models or, maybe, as theories, depending on their overall qualities. Hence, only theories of type IV – for explanation and prediction, are from Weber's point of view appropriately called theories. Huxham and Sumner (2000) contribute a model in which observations and hypotheses may evolve into models as they gain explanatory power and in which theories may be broken down into falsifiable laws and facts. Models and theories provide the greatest explanatory power and theories differ from models in that they have greater scientific status. As such, theories and models are essentially the same thing, similar products merely at different stages in a process of scientific refinement. The present view aligns with these notions and with the view that the product and the process of theorizing are different but dependent, thereby acknowledging the view that the product cannot be understood and used properly without sufficient knowledge about the process from which the product came to be. Hence, proper notions for theoretical contributions, that is, whether to call them typologies, models, theories, or frameworks, are here considered depending on the nature of the problem, the explored phenomenon, and on the character of the theorizing process.

Ultimately, the main problem seems to be the problem of the problem, that is, the problem of identifying what practice or real world phenomenon that needs to be influenced through enhanced understanding. Identifying relevant problems require, arguably, a critical and open mind with insights from and knowledge about the problem domain, as well as an analytical and conceptualizing mind in order to formulate plausible explanations of significant problem aspects and effects. It seems that for relevance to practice or practitioners, of theoretical contributions, it is required that scientists generate research questions through problematization (Sandberg and Alvesson 2011, Alvesson and Sandberg 2011) or construct theories through disciplined imagination (Weick 1989) about relevant problems. Again, the question is, how are relevant problems identified? Corley and Gioia (2011, p. 13) define *prescience* as "the process of discerning or anticipating what we need to know and, equally important, of influencing the intellectual framing and dialog about what we need to know". Practice oriented researchers and research oriented practitioners ought thereby be highly valuable when striving for practical utility and relevance to practice because it appears likely that prescience emerge within the borderlands between practice and science.

This research aims for a high level of relevance to practice, based on the nature of the studied phenomenon. The phenomenon studied is usage of advanced, computerized, technological systems and how practice is influenced by the contemporary view of system usefulness, which is a phenomenon with far-reaching implications for practice because today human-(computerized)-technology relations influence just about everything we do. The nature of the identified problem is thereby of a kind covering a wide range of aspects, including undesired social consequences of an inappropriate view of technological usefulness. As such, the present research result, the theoretical contribution, the tools for enhanced understanding, are unlikely to fit the notion of predictive theory. The kind of understanding required to influence practice in this context is more aligned with analysis and explanation, thus better represented by terms such as typologies and models. However, theories come from models and predictive theories require, arguably, typologies and frameworks for a comprehensive understanding of their applicability and, thus, the present theoretical contribution may very well constitute a basis for development of theories as the result of future research. In addition, because this research focus on theorizing on a rather high level of abstraction, the developed typologies and models are complemented with conceptual frameworks further supporting a comprehensive understanding. To summarize, phase one, designated the theory generation phase, does not present predictive theories, but it is still considered a theoretical contribution.

Prescience (ch. 2.2.4), seems crucially important for generating theory relevant for practice and, as a consequence, it is important to convey as

thorough as possible (ch. 2.2.5) the nature of my particular prescience as it is significant for assessing the relevance to practice (ch. 2.2.6) of the theoretical contribution made by this research. The nature of my prescience is naturally shaped by the nature of my particular experiences and preconceptions, which is why selected aspects of my professional background are recapitulated in chapter 3.

2.2.4 Prescience

In short, prescience comes from having experiences and preconceptions that combined with insights – both scientific and practical insights – facilitates the development of well-spotted research questions and theories, in its broadest sense, as well as relevant questions and theoretical contributions endowed with a rather extensive set of disparate qualities, if one accepts Weick's notion of good theory, that is. (This notion was stated in the beginning of ch. 2.2.2). Quite a mouthful, is it not? At the same time, while pursuing abstract research questions implying a strong focus on theory development, with critical realist philosophical assumptions relying on retroduction that implicitly assume good analytical qualities, the previous claim (in ch. 2.2.3) that this research aims for a high level of relevance to practice needs to be followed up. The purpose with the following discussions about prescience and explication of experiences and preconceptions (ch. 2.2.4 and 2.2.5) is therefore to provide plausible grounds for relevance to practice (ch. 2.2.6). This purpose forces me to elaborate on the quality of my own prescience, which feels highly awkward because it is likely to give an impression of self-justification. It appears, however, necessary because to ground means, "to provide a basis or reason for (something)".[17] Hence, I feel strongly obliged to disclaim hubris.

The present research is associated with what ought to be a pleasant problem, the problem of how to relate to the supposedly relevant and rather exclusive knowledge facilitated by my professional experiences. As a researcher, if I had been considering to do field studies, if I had wanted to gain a rich understanding of usage and usefulness of advanced technological systems in uncertain and demanding situations, then I assume that fighter pilots and military aircraft systems would have been highly significant subjects. That is, the present professional me do find someone like the former professional me an interesting individual to consult and interview, a research situation for which the problem mainly is methodological.

Besides clearly facilitating having experiences from the borderlands between practice and science, the upside of my double position is that, arguably, I have no problem in understanding my own view of things. It is a position that effectively removes one of the major difficulties in interpretive

[17]http://www.merriam-webster.com/dictionary/ground

research, the problem of assessing whether made interpretations are valid from the perspective of the studied subjects (e.g., Walsham 1995). From the research perspective, this means that my personal insights and experiences constitute a rich understanding supposedly of unusual validity if viewed from the perspective of other subjects from the field. The downside of this double position is that it makes me highly susceptible to two archetypal mistakes for qualitative researchers, which are having an elite bias and the risk of going native (Miles and Huberman 1994, p. 263). By elite bias, it is meant the tendency to give too much credit to some informants and disregard others, and it seems likely that I would consider my own thoughts overly representative, while it is undeniable that my personal view is merely one view. It is a view that cannot be taken as general for users of advanced technologies in demanding environments, not even as the view of fighter pilots in general. By going native, it is meant the risk of losing the research perspective when becoming too much involved in the studied environment, and I was once fully native.

On the other hand, this research is not strictly interpretive, making the methodological problem not be to address specifically the elite bias and going native problems. Rather, the methodological problem is how to convey explicitly information about experiences and preconceptions, information that might have been considered showing an elite bias and indicating the problem of going native for interpretive research. The methodological challenge is to convey this information such that implications for theorizing and retroduction become comprehensible. However, in order to provide as thorough grounds as possible for prescience some notions from interpretive research will be used that, as it seems, in relation to my particular background to some extent also neutralizes the problems of elite bias and going native. These grounds and the character of my particular prescience will be returned to when discussing relevance to practice (ch. 2.2.6). First, however, let us consider the problem of conveying information about experiences and preconceptions more generally.

2.2.5 Making experiences and preconceptions explicit

Making practical experiences explicit and conveyable is a true challenge because much of it belongs presumably to the tacit dimension of knowledge, a dimension characterized by Polanyi (1966) as intrinsically different from the explicit dimension, yet inescapably coupled. Tacit knowing consists of two terms, Polanyi suggests, the functional structure and the phenomenal structure, and “neither is ever present without the other” (1966, p. 7). The functional structure of knowledge is about knowing-that and the phenomenal structure is about knowing-how (Ryle 1945). Polanyi continues by defining two additional aspects of knowing, the semantic aspect that denotes the meaning of what is known and the ontological aspect that denotes what the

knowledge is about. Dreyfus & Dreyfus (e.g., 1986, 1988) refer to both Ryle's knowing-that and knowing-how, as well as Polanyi's tacit and explicit dimensions, but they choose to introduce yet another distinction, that between a detached or an involved perspective. The detached perspective is here interpreted as covering knowing-that and certain explicit semantic aspects (cf. knowledge and intellect in 2.1.4), while the involved perspective covers knowing-how and tacit semantics (intuitive meanings).

On the other hand, there is for example Nonaka (e.g., 2000, p. 9) who suggests that "when tacit knowledge is made explicit, knowledge is crystallised, thus allowing it to be shared by others, and it becomes the basis for new knowledge". The present stance is, however, principally the former, that of Polanyi, Ryle and Dreyfus & Dreyfus, which is to assume that truly tacit knowledge exists that therefore is impossible to describe explicitly to its full value. Especially, certain rich meanings and, perhaps particularly certain aspects of knowing-how that in turn may be connected with Aristotelian practical wisdom, are considered impossible to make explicit. In some sense, the relation between tacit and explicit knowledge may be described as having the same kind of relation as that between critical realist intransitive objects and transitive objects of knowledge. While it is possible to produce descriptions of tacit knowledge, the explicit transitive description will never be equal to the real intransitive object. To continue the critical realist philosophy of science metaphor, this difficult relation does not mean that tacit knowledge is of no value to science, it does only mean that scientific work face a never ending challenge to make appropriate use of the transitive objects of knowledge about tacit intransitive objects.

The strategy adopted here is to strive for some kind of shared understanding of relevant and presumably to great extent tacit objects of knowledge from my background, by providing narrative descriptions intended to indicate the kind of experience that is assumed relevant for theorizing about situated usefulness. These descriptions are deliberately tried to be held brief (with varying success), in order to avoid too lengthy stories that may obstruct the object of knowledge by unintentional deception towards a counterproductive focus on insignificant details. This approach presupposes, however, readers to have experiences of their own. Experiences sufficiently similar in character to my experiences, for an ability to imagine the character of the tacit object described by the narrative (e.g., without experiences from flying, experiences from car driving might suffice for getting the essential picture). To support further such imagination, efforts are therefore made, as far as possible, to include everyday examples, perhaps trivial, that nevertheless are similar in a principal sense to the described objects of experience (making some narratives not so brief after all). For the present purpose, one period or aspect of my background is selected and narrated for each of the main chapters presenting the frameworks that constitute the theoretical contribution. These narratives are presented in

chapter 3 – Experiences and preconceptions. How the narratives relate to the frameworks and the different chapters is described further in chapter 2.3 Research method.

2.2.6 Relevance to practice

The problems of elite bias and going native associated with interpretive field research depend, arguably, on the assumption that theorists cannot be sources of empirical data themselves and that sources of empirical data (i.e., members of the field) cannot be theorists. In terms of classic grounded theory (GT) theorizing, Glaser (2002, p. 29) holds that when taking this shortcut of using theories from participants, researchers "forget that participants are the data, NOT the theorists". In my view, that stance is to reify the categorical distinction between science and practice (discussed in ch. 2.2.2). Moreover, the discussion above (ch. 2.2.4 & 2.2.5) suggests that valuable prescience develop within these particular borderlands, thereby implicitly making prescience to some extent require being both data and theorist, at least interchangeably. George and Bennett (2005) stress that iteration between theory and data is essential for theory development in the social sciences. What will be elaborated on in much more detail in chapter 3, the rather special structure of my earlier career has made me alter continuously back and forth between being the data and the theorist, if seen from a GT perspective. Over a period of about twenty years, I have shifted between working practically as a fighter pilot and scientifically with matters related to this practice as a pilot engineer, all depending on which hat (or helmet) I had on. For a pilot engineer, the swapping of hats could be for a few hours of a particular day, for entire days, or for weeks or longer. For a Swedish Air Force pilot engineer the distinction between data and theorist is blurred on purpose, in order to reach insights from both perspectives (see ch. 3.2). As I see it, the character of my background has allowed for some testing of my experience-based theories on fellow field subjects and the structure of my career has rather implied going in the opposite direction than towards becoming native, thus counteracting both problems mentioned. Once I was fully native, now I am a fulltime theorist, although with a certain amount of presumably rather valuable experiences and insights about real issues, thereby, supposedly, having the ability to choose relevant problems leading to theories relevant to practice.

Mainly the practical part of my background is covered in chapter 3, although some interpretations indicating underlying theorizing are included as well. For further understanding of my conceptualizations, each chapter in phase one (ch. 5 - 9, and ch. 10 that builds upon the others) include also a recapitulation of my understanding of related theoretical areas. In conjunction, these external theories and experience recapitulations aim to constitute sufficient descriptions of aspects relevant for assessing my

prescience thereby indicating the kind of relevance to practice the theoretical contribution may have.

2.3 Research method

The present research is not strictly inductive, partly because the critical realist aim for explanations of underlying mechanisms require other methods than purely empiricist ones. This research cannot claim objectivity and rely on positivist hypothetic-deductive methods because the developed explanations depend too much on subjective interpretations, and because of the aims to include aspects not suitable for explanations from a detached and objective perspective. Nor can this research pride itself with having utilized a multiplicity of diverse critical naturalist (critical realist) methods that by triangulation provide rigorous validity for developed explanations. Yet this research may be considered to include some of all the above.

To begin with, the overall method includes the three phases that were introduced in the thesis structure (ch. 1.2.3). The first phase, exploration of usefulness and theory generation, represents in some sense the context of discovery, in relation to the following two phases that align more with the context of justification. This distinction is, however, not always completely clear. Hoyningen-Huene identifies no less than five different distinctions between these two contexts and suggests one that may cover them all, a distinction between two perspectives: "an abstract distinction between the factual on the one hand, and the normative or evaluative on the other hand" (2006, p. 128). The notion of research phases may give the impression of a separation in time, the most straightforward alternative of the five distinctions by Hoyningen-Huene. However, phase one relies also on elements of knowledge about consequences and norms that are essential also for the context of justification. Hence, phase one is research within both contexts thus relying on the abstract distinction, although with emphasis on the context of discovery. The situation is similar for phase two. Focusing on empirical corroboration of generated theories makes phase two have a clear emphasis on the context of justification while also providing factual input to the exploration and theory-generation process, thus working in the context of discovery as well. Phase three continues within the context of justification, by further discussing plausibility of aspects and by clarifying the overall line of reasoning. The normative and evaluative perspective is elaborated on in phase three, in terms of normative considerations, as a discussion about research implications, and as philosophical reflections.

The overall research method, especially when including the case studies in phase two, may be described as "integrating comparative and within-case analysis", with the aim to develop typologies or typological theories (George and Bennett 2005, chap. 11). As such, this research comprises elements of

both empiricist and hypotetic-deductive methodological approaches (i.e., drawing conclusions and generating theories from case study data as well as testing hypotheses – generated/hypothesized theories – on real world cases). Phase one focus on the first three methodological principles of critical realism (ch. 2.2.1), phase two on the first four, while the fifth principle in a sense is addressed by the phase structure because the distinction between phases may be viewed as implying a kind of methodological triangulation. Typologies and frameworks for describing and analyzing the studied phenomenon (situated usefulness) are generated in phase one by use of one method, then the object of research is explored and the theoretical contribution is corroborated in phase two by another method.

2.3.1 Phase one – theory generation

The overall purpose of this research is to scrutinize the contemporary view of usefulness and the main objective is to contribute an alternative definition. Usefulness of technological systems is the phenomenon of concern and, as was described above, the methodological approach in phase one is focusing on the first three principles of critical realist methodology. They were; first, to explicate events (i.e., to describe aspects relevant for technological usefulness); second, to explicate structures and contexts; and third, by retroduction identify and elaborate on underlying mechanisms plausibly generating the studied phenomenon. The expected result is a set of models and frameworks (typologies) explaining identified structures (e.g., worldview descriptions), powers (e.g., system property descriptions), generative mechanisms (e.g., relations and tensions), and tendencies (e.g., cultural trends and influences).

In order to understand its nature as well as what makes a technological system truly useful, the concept of usefulness is broken down recursively through problematization (Alvesson and Sandberg 2011, Sandberg and Alvesson 2011). It is broken down into related concepts and areas of interest by use of the what, when, where, who, why, and how questions (ch. 1.2.3) that are considered the building blocks of theory (Whetten 1989). This problematization process is also reflected in the thesis chapter structure, which is elaborated on below in connection with Figure 2.3. Finding answers to these questions may be described as an abduction process that according to Charles S. Pierce is a kind of reasoning that "lead to judgments about *pursuitworthiness* of theories" (McKaughan 2008, p. 446, emphasis in original). Abduction is closely related to retroduction and both concepts describe a process that is inductive and deductive at the same time. "In many research projects and research programs, a combination of induction and deduction is useful or even necessary, depending upon the research objective, … and availability of relevant cases to study" (George and Bennett 2005, p. 239). Phase one is essentially about pursuing RQ1 (what is

the nature of usefulness?) and for clarity and concreteness the case of use is taken as the starting point. Hence, phase one does also pursue RQ2 (what is the role of the human user and how does this role affect actual usefulness?). As such, phase one is a parallel inductive and deductive process that in the present research case largely is governed by personal experiences and preconceptions (i.e., prescience).

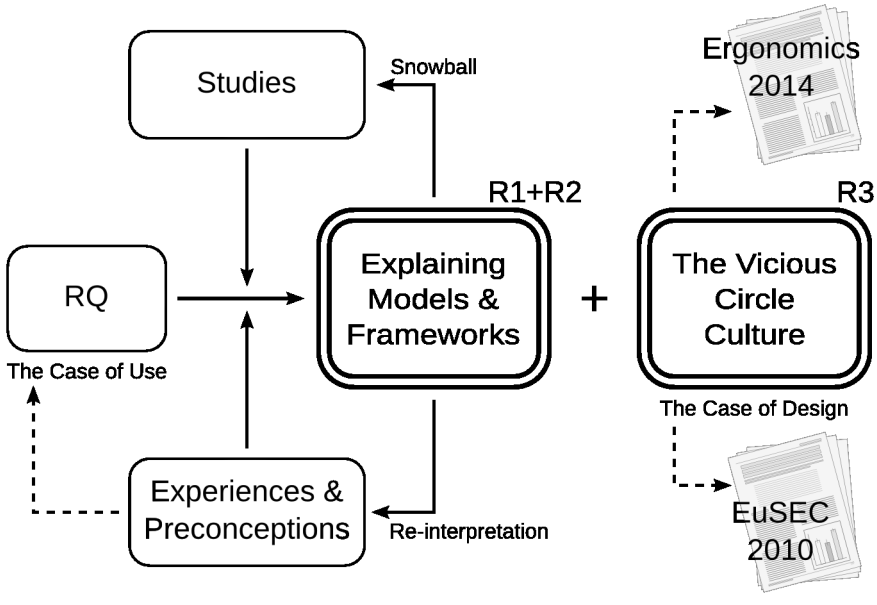


Figure 2.2: Conceptual design of research – phase one

The inductive and deductive theorizing process of phase one is illustrated by Figure 2.2. For the deductive loop (Studies → Models → Studies) a snowball approach for choosing topics is selected. New theories, areas of interest, and literature to study are chosen based on pursuitworthiness judged by previously gained insights and theoretical knowledge. The set of theoretical topics studied as a result of this snowball-selection approach includes philosophy, systems theory and systems thinking, systems analysis techniques, systems engineering, human factors and safety research, as well as several topics related to psychology such as human behavior, human error (but, naturally, also studies of human strengths), mental properties, and decision making. These areas are briefly recapitulated in the chapters constituting the body of phase one (i.e., ch. 5 - 9). For the inductive loop (Experience → Models → Experience), the methodological principles and the problematization approach described above are essential (ch. 2.2). The inductive loop represents a process leading to continuously enhanced models and frameworks. Overall, these two processes support each other mutually.

Theoretical input is grounded and understood in terms of experiences and preconceptions while preconceptions are adjusted and the understanding is enhanced by further theoretical studies.

Generated theories (in Figure 2.2, the double-lined box: Explaining Models & Frameworks) aiming to explain the nature of usefulness is the first result of phase one, accordingly designated R1. The second result, R2, is the contributed redefinition of usefulness. R2 meets the main objective by answering RQ1. R1 fulfills the overall purpose and answers RQ2, to some extent at least. While phase one focused on the case of use, another result is mainly related to the case of design thus providing grounds for answering RQ3. The third outcome from scrutinizing usefulness is the identification of a tendency, a socially embedded mechanism that seems a plausible explanation for why some systems become designed such that they do not become as useful as they were intended to be. This tendency is a result labeled R3, identified through the above described theorizing process in combination with the empirical corroboration process in phase two (the case studies). The subject of R3, rhetorically designated *the vicious circle culture* (elaborated on in ch. 15 – Research implications), has also led to two publications (Stensson 2010, Stensson and Jansson 2014a).

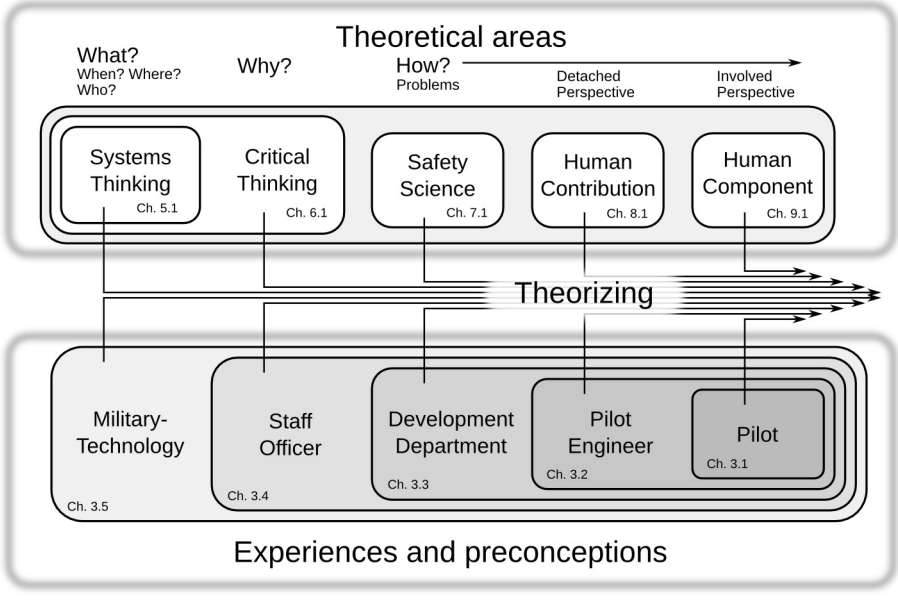


Figure 2.3: The relation between theoretical areas and described experiences

Figure 2.3 aims to further clarify the retroduction and theorizing process, as a methodological effort to support the plausibility of retroduced aspects of usefulness (see later also ch. 16, especially ch. 16.4 Validity of contributed

models and frameworks). All the theoretical areas (the recapitulated external theories) and all described experiences have, naturally, influenced the entire retroduction and theorizing process. That is why the set of theoretical areas (the upper half) and the different kinds of experiences (the bottom half) are depicted in Figure 2.3 within one uniting box each. However, the different areas of experience are considered more interwoven, in the sense that they are more difficult to separate from each other, compared to the different areas of scientific knowledge. Experiences are, arguably, difficult to separate from each other much because of their tacit nature. Experiences are also believed to develop recursively over time, most often starting with the obtaining of certain skills or with the learning of specific pieces of knowledge, later in life expanding to aspects that are more general, through reflection about implications and through identification of holistic effects. With that said, it should be clear that experiences tend to apply to problematization in the opposite direction compared to how they develop.

The process of exploring usefulness by problematization illustrated by Figure 2.3 connects to critical realist methodological principles as follows. Explication of usefulness as a phenomenon is pursued by asking the questions, what, when, where, and who, in a holistic sense. Explication of structures and contexts is initially pursued by scrutinizing critically why technological systems are useful. The explication of the phenomenon, of structures and contexts, is then continued by asking how technological systems become useful, and by specifically asking about the human role in this matter. For retroduction to plausible explanations, both theoretical knowledge and practical experiences are relevant, and for assessing the quality of retroduced explanations these foundations should be as transparent as possible. The main purpose of Figure 2.3 is to support such transparency. To support this transparency further, the different theoretical areas are recapitulated as theoretical frames of reference throughout the chapters of phase one. The purpose of chapter 3 is to provide means for some kind of understanding of my experiences and preconceptions.

2.3.2 Phase two – case studies

The overall purpose of phase two is, naturally, the same as for phase one, but in this case with an additional focus on empirical corroboration. In order to create a further enriched understanding of situated usefulness a qualitative analysis was performed, corresponding to the first two principles of CR methodology: explication of events, and explication of structures and contexts. The main components of qualitative analysis are data reduction, data display, and conclusion drawing/verification (Huberman and Miles 1983, Miles and Huberman 1984, 1994), where the latter corresponds principally to the retroduction and empirical corroboration principles. In this context, verification takes essentially the form of a qualitative validation of

the theoretical contribution (i.e., empirical corroboration), by the providing of empirical grounds for assessing the plausibility of phase one results. The three components of qualitative data analysis are visible in the arrangement of the principal workflow illustrated by Figure 2.4.

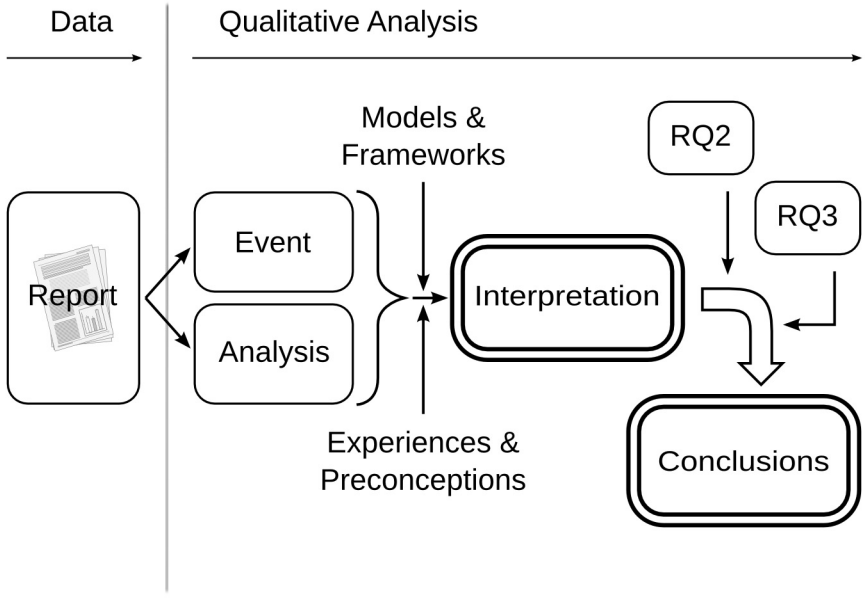


Figure 2.4: Conceptual design of research – phase two

The raw empirical data is available as official reports describing the four accidents and the three future technology usage scenarios. First, the reports are studied and recapitulated, split into an explication of the event and a description of the analysis made by the issuing authority, the latter essentially working as a tentative explication of structures and contexts as well as a tentative retroduction to plausible powers and mechanisms, commonly phrased as causes for events reported by incident investigation authorities. Such a data processing effort implies both qualitative data reduction and data display. Second, a qualitative analysis was made by applying the theoretical contribution from phase one, as well as personal experiences, preconceptions, and insights, the latter perhaps particularly for the airline cases. The result of this analysis is designated interpretation in Figure 2.4. Such an analysis effort implies further data reduction and display efforts as well as retroduction and conclusion drawing. Third, as a last step, overall conclusions were drawn from the analyzed data of all cases in conjunction, about situated usefulness. Because this step is made in light of RQ2 and RQ3, the results provide grounds for answering both.

Whereas Figure 2.4 may give the impression of a sequential applying of analysis components, the above description of the overall method and the method for phase one should make it clear that in practice they have been made in parallel. Furthermore, the figure indicates that the research questions (RQ2 & RQ3) were considered only in the last step, common for all cases. However, they are implicit also in models, frameworks, and typologies, as well as in the application of experience-based knowledge used to produce the interpretations.

2.3.3 Phase three – discussions

It is somewhat irrelevant to talk about a specific method for phase three, as it is more of a summing up discussion. However, the third phase may be considered to represent an essential part of the empirical corroboration methodological principle, to analytically validate proposed mechanisms and assess their explanatory power. After discussing implications of the contemporary view of usefulness in terms of the (for human emancipation) assessed counterproductive shift in meanings for essential concepts, perhaps particularly for the concept of autonomy (ch. 15), a logical structure for plausibility of retroduced theories and frameworks is presented (ch. 16). This approach aligns with CR philosophy because retroduction involves both production of plausible explanations and reasons for why they are plausible. The theoretical contributions are sorted into different categories, in an explicit effort to clarify how things relate to each other. Some contributed models are intended to work as explanations for different aspects of usefulness, others explain why the explanations are plausible, and yet another set of frameworks function as foundations for understanding both the explanations of usefulness and the reasons for their plausibility.

3 Experiences and preconceptions

The purpose of this chapter is to give the reader as much insight as possible into certain experiences and preconceptions from my background considered relevant in this research context, as a foundation for prescience and as the rationale for choosing the research questions. As such, this account works as grounds for assessing the relevance for practice and practitioners of the theoretical contribution. For this purpose, it appears necessary once again to point out the suggested relation between my particular area of experiences and a general understanding of human-systems interaction (introduced in ch. 1.3.2). While the case of military aviation might be considered an extreme special case thereby not actually relevant when studying everyday gadgets, I maintain that many issues are principally the same for all kinds of technological systems. The difference is that consequences are much more visible, issues are much easier to discern, and consequences of inappropriate designs develop much quicker within extreme situations, which makes fighter aircraft perfect systems for scrutiny. That is, the extreme case is not irrelevant but instead what might open our eyes to issues that otherwise may evolve slowly and become great problems in the future. Hence, the extreme should be considered benign breeding grounds for prescience, should it not?

What follows is a purposefully selected set of narratives about not entirely lifelike stages of my career, but stages conceptually adjusted to suit the structure of this thesis as was illustrated by Figure 2.3. The purpose of the following five-chapters-long life-story is not to provide a batch of refutable facts. The purpose is merely to provide cues for a richer understanding of supposed mostly tacit areas of knowledge. Presumably, the kind of understanding aimed for requires therefore that readers actively try to relate the cued experiences to own experiences of similar kinds. The five sub-chapters address in turn the following areas of knowledge:

- First-hand experience of knowing-how, of issues related to edge awareness, and of the generative dimension of system (i.e., vehicle) control, presented in light of Swedish fighter pilot training.
- Analyses of (tacit) knowing-how, of the human role in relation to technological systems, in light of first-hand experience from when things are about to go wrong.
- The involved perspective of human factors in aviation, scrutinized in terms of human decision-making and some paradoxical issues for

traffic safety. The two-modes-model of human thinking (the intuitive system 1 and the rational system 2) is complemented with a notion of a semi-ration system 1½, discussed further later in the thesis.

- The external (detached) view of the human role in relation to technological systems, phrased in terms of desired end-states as well as commander and system-operator relations. Tensions between purposes and the use of models are considered by discussing simulators and effects of simulator training.
- Philosophy, by briefly discussing the multidisciplinary domain of military science and explicating the subtle but crucial difference between studies of military-technology (one word) and studies of military technology (two words).

Personal experiences develop, arguably, to a great extent recursively, beginning with concrete specifics and gradually expanding to include abstract generalities. This development comes about, naturally, by virtue of a continued collection of experiences from explicit and specific work related events, but arguably also by age and through the gaining of life experiences in general. The theoretical frameworks in phase one, on the other hand, are arranged in the reverse order, beginning with the general worldview and ending with specifics about (individual) human-systems interaction. There is, arguably, a touch of natural logic about that ordering as well, as it is obvious that the entire set of experiences have been influential for all sets of frameworks and models during this research. The description of my work-life experiences, however, are best presented in chronological order starting with the specifics, as the story, so to speak, is cumulative. It seems therefore just as well to start at the very beginning.

As the oldest sibling and a boy born in the late 1960s, it was completely normal that I came to steer the family car sitting in my father's lap as soon as I was tall enough to see above the dashboard. During harvest each fall, my favorite position was at the wheel of my grandfather's farm tractor, a position I was encouraged to occupy long before I was able to press down the clutch. One could say that I was early conditioned to enjoy driving vehicles. There was also an early interest in computers triggered when we provided ourselves with one of the early home computers in the late 1970s, resulting in that I became hooked on programming, as a hobby, but with a quite intense peak during my Master of Science studies. However, my overall interest in technology was for a while, perhaps because of conditioning, rather focused on things with engines that you could drive, and one of my first real jobs was accordingly as a car-courier driver. When the opportunity then appeared during my compulsory military service to try for becoming a Swedish Air Force pilot with the prospect of being allowed to fly military fighter jets, from a driver's perspective perhaps the most impressive kind of motorized vehicles there is, I obviously went for it. To my extreme surprise,

I actually passed all the tests and I was accepted into flight school. This was about twenty-five years ago.

3.1 Training for edge awareness

There are a few things worth mentioning in this context about fighter pilot training, and perhaps particularly the Swedish way of doing this. To begin with, being a pilot requires a lot of practical training because flying is largely a craftsmanship. Flying is, as any other activity primarily based on skills, something you have to learn much by trial and error because it is impossible to learn the essential aspects (i.e., the necessary skills) from reading a book. You simply have to build personal embodied experience in order to become comfortable with the task, as the first step towards proficiency and expertise (e.g., Dreyfus and Dreyfus 1980). It is likewise impossible to learn to ride a bicycle solely from theory. The trick is to harness the errors required for gaining experience such that you can learn from them without suffering a capital loss, which unfortunately is a rather common scenario in the pilot business. While learning to bike you might earn yourself a few bruises and scratch marks, but it is unlikely that you will crash and die as when learning to fly. Still, experiences from close encounters are, arguably, required for knowing the limits of the aircraft and yourself, and thereby required for an ability to sense upcoming hazards. The above includes in fact a training paradox well captured in the following quite obnoxious words of wisdom: – While an unerring judgment comes from comprehensive experience, it is an erring judgment that makes you experienced! It seems we human beings can only learn things superficially from instructions and artificial training environments because it appears that we at large must experience things first hand in real situations, in order to understand truly.

In military flight schools, this kind of trial and error education is largely managed by setting up significant situations with trainer aircraft allowing you to experience the effects of an erring judgment, but with the backup of an experienced instructor pilot able to help you recover until you can manage for yourself. There are (at least) three crucial aspects of this kind of training relevant for the present context. First, you train explicitly to recognize and cross the boundaries for (i.e., the edges of) safe flight, which are a set of situation-dependent limits making up what is called the flight envelope of that particular aircraft.[18] Second, you train explicitly to recover from situations you arguably should avoid ending up in, by deliberately and

[18]The flight envelope consists primarily of maximum altitude, and minimum and maximum speed for leveled flight at different altitudes, which are parameters that in turn depend on environmental conditions such as temperature and air pressure. For maneuvering aircraft (like fighters) the flight envelope is further complicated because dynamic aspects add several more limits and interrelated parameters to the situation.

repeatedly create such situations and practice recovering until you manage satisfactory. This kind of education seems very much in contrast with what happens in the civilian society today, in which the common attitude is that you should not practice recovering from forbidden situations because it implies that you must create such forbidden situations, which, of course, from a formal perspective is forbidden. It is instead considered safer to make sure (commonly by technological means) that these forbidden situations never occur, which thereby, in theory, render training recovering from such situations obsolete. The Air France disaster discussed at length in chapter 12.3 is a tragic proof of the inadequacy of this calculative approach. Third, you do this kind of training in real aircraft in real airspace thus creating real-life experiences, as opposed to the slightly superficial, somewhat theoretical, and artificial experiences gained in simulators (discussed further below, ch. 3.4). I have now come to consider much of the training in flight school to be an explicit education towards edge awareness for aircraft maneuvering, while that exact phrase was never used. One phrase that sometimes could be heard was 'learning to fly by the seat of your pants', suggesting getting intuitively familiar with the maneuvering task thereby leaving (rational) mental capacity to maintain good airmanship, which is another concept somewhat related to edge awareness. Airmanship is a term with broader connotations that includes aspects relevant also for being a trustworthy captain of an air vessel. Another obvious phrase was simply: 'learn to stay on top of things', which have a connotation both of maintaining airmanship and intuitive maneuvering skills. This last phrase is, arguably, an essential aspect of the human factors in aviation buzzword, situation awareness, a concept elaborated on in chapter 9.

The Swedish way of training military pilots is allegedly somewhat special compared to what appears as standard international praxis. In Sweden, the strategy is to rely heavily on a thorough screening process with an extended testing procedure aiming to minimize dismissals during education. This strategy has one crucial and highly beneficial consequence, besides that of being cost effective.[19] The approach creates a supporting educational atmosphere, as opposed to the competitive culture that tends to spring from a screening strategy based on large intakes and extensive dismissals. Within a competitive environment, fellow prospects become opponents and instructors tend to take on the role of inspectors rather than coaching educators, thus encouraging a cover-up mentality effectively inhibiting a shared gaining of experiences. On the other hand, as a direct result of the more supporting culture it becomes easy to learn from the mistakes of other prospects and to willingly accept valuable guidance from instructor pilots with whom problems and mistakes can be discussed at ease. One tangible

[19]Flight training tend to be rather expensive, and to begin training people that later are to be dismissed is therefore a genuine waste of resources, not to mention the destructive aspects of personal failures (i.e., the psychologically trying experience) associated with being dismissed.

result of this cooperative culture is a large database maintained by the Swedish Armed Forces Flight-Safety Department, a database of incident reports. These reports are called 'disturbance notes',[20] covering everything from small mishaps and mistakes to severe accidents. They are issued by the pilots themselves[21] and because of the friendly atmosphere, they are assumed openhearted and honest. During the daily debriefings at the squadron, reported disturbances are always discussed and analyzed, with the prime purpose of increasing the shared grounds of experiences.

Flying, bicycling, and car driving, these are all activities that include low-level interaction with a technological system in a manner that require bodily skills as well as high-level intellectual (analytical and rational) thinking. These activities are also taking place in a dynamic physical world with real and possibly lethal hazards. Cars and aircraft, but perhaps not so much the bicycle (I am thereby deliberately disregarding the increasing amount of small gadgets and smart-phone applications emerging also for bicycling), are also systems that are getting more and more computerized. Vehicle computerization has up until recent primarily taken the form of supplementary systems involved mainly in high-level decisions (e.g., navigation, communication, etc.), but today systems are implemented that take active part also in actual vehicle maneuvering. For example, traction control and automatic parking systems, as well as fly-by-wire flight control systems. One essential aspect of this story is the importance of the two-way connection between intuitive skillful maneuvering and deliberate rational decision-making. The character of your bodily skills, your awareness about them, and your confidence in them have great influence on the decisions you make, which makes first-hand experience of such low-level skills of crucial importance. This is what flight-school and tactical training is all about. It is about gaining experience from handling the performance edges of your aircraft. The purpose is to learn to trust your own skills as well as trusting the skills of your colleagues, which, for example, is necessary when flying in formation as a subordinate number, especially in bad weather.

For a while, I had the opportunity to exercise my edge awareness fulltime, together with colleges having a similar mindset. We did this will all vehicles at our disposal. There was some skidding with the squadron van, for which opportunities were many because winters are long at the most northern Swedish fighter wing. We practiced aerobatics with our small propeller aircraft (SK61) and the twin-jet trainer-aircraft (SK60). Mostly, however, we practiced for our position as tactical reconnaissance pilots, flying the reconnaissance version of the Viggen fighter (S37) in air-to-air combat and dogfight maneuvering, and in a lot of high-speed tree-top-skimming missions over the scarcely populated expanses of northern Sweden.

[20]My own guess for a translation, it is in Swedish called “Driftstörningsanmälan (DA)”

[21]There are others besides the pilots also issuing disturbance notes, such as maintenance personnel (aircraft mechanics) and (military) air traffic controllers and meteorologists.

3.2 Pilot engineer

Because of my personality, my genuine interest in computers, or more probable, from a combination of these and other factors, rather quickly I became interested in what at the time most often was referred to as Man-Machine Interface (MMI) issues. I realized that my strong interest in driving vehicles largely was fueled by the fascination of being able to control these systems. Consequently, my interest homed in on human-systems interaction issues and aspects about how beneficial system control is possible. This interest applied, to begin with, mainly to the quality (or lack thereof) of the increasing amount of computerized human-machine interfaces being developed for all kinds of systems, not least in military aircraft prospects. This interest became strong enough to make me apply for an alternative military career, in Sweden called pilot engineer. The idea from the Air Force perspective was then that when the JAS39 Gripen, under development at the time, was to enter into service, each squadron was to have at least one pilot engineer to assist in coping with the comparatively much more computerized aircraft. It did not work out exactly that way for me since I never returned to work fulltime at the squadron. What happened was that I moved to Uppsala and studied for five years at the university, towards a Master of Science degree in engineering physics, specializing in computer systems, although I continued to fly at the northern squadron whenever the study schedule and squadron activities permitted.

The first summer, after an intense first year at the university implying the longest period so far in my pilot career without any flying whatsoever, I learned an important lesson about the earlier mentioned two-way connection between intuitive skills and rational decisions. First hand, I experienced the fact that your edge awareness (your intuitive sensing of performance edges) decays quicker than your rational knowledge about this fact. Actually, I thought I knew about this even before, and supposedly, most people feel the same way. However, the crucial question is whether most people truly know about this? Arguably, you do not really understand how it feels to burn your hand on the stove until you actually have done it, no matter how much you are warned. One can hope that you only have to burn yourself slightly to reach sufficient insight. To understand truly how it feels when your intuitive knowledge about your edge awareness differs from your actual edge awareness, supposedly, you must experience it in a way that makes the experience stick. For me, the following experience gained in a Viggen fighter is one that seems to have remained tangible. The story is as follows.

Before the university, I had experienced a continuous increase in flying skills with never more than a few weeks of vacation without practice. I had learned to 'fly by the seat of my pants' and I had learned to trust my ability to do so. This fine summer day, when I finally came to steer a Viggen again, after what seemed so long, I did what I used to do. During initial warm-up

aerobatics, after making sure I once again had enough speed to complete it easily, I entered into a looping by pulling the stick backwards to catch the maximum limit of 6 Gs. On the way up through the sky I just looked out through the canopy, enjoying myself immensely and the beautiful view, while keeping the G-load 'by the seat of my pants' as I had been able to do with great precision before. However, the 'skill of my pants' and my earlier quite well calibrated intuitive sense of speed had evidently decayed. Without sensing it I relaxed the Gs, making the aircraft turn more slowly and the looping became too wide, resulting in a higher top altitude for which I did not have the speed nor the thrust (in the beginning of a sortie the aircraft is very heavy with fuel). The situation turned into one that merits the phrase every flying story begins with, namely: 'And there I was...'

And there I was, less than halfway through the looping with the nose still pointing to the sky but without speed. I was in for a rude awakening. The joyful and relaxed situation became instantly deadly serious. Fortunately, the Viggen was a well-built aircraft with rather intuitive aerodynamics. What happened then was that since I already was inverted (second quadrant of the looping), I was able to keep the aircraft for a while in the aerodynamically desirable position having the fin pointing to the ground, keeping it within undisturbed airflow to maintain longitudinal stability, aiming to fall through the top of the failed looping, now virtually without speed. At some point near the top, however, the engine pumped due to compressor surge, a likely consequence of too low speed and an unstable airflow through the inlets, which required me to pull the throttle back to idle. Without thrust, the situation became even worse. Before I managed to complete falling through the top of the looping and gain speed again from diving, the aircraft flipped around. Evidently, I had lost control of it. However, a moment later I managed to regain control and get the nose pointing down, dive for speed, level out, check that the engine still was running properly and enter into a controlled and level flight at ease. Looking to get on top of things again, I circled my allotted practice airspace for a while before returning to base authoring a disturbance note.

This event was, with hindsight, an extremely valuable lesson, although a rather intense sensation while experiencing it. Everyone having been part of, or sufficiently close to become part of, a severe accident while driving would presumably sympathize with this feeling. Regardless, it was a rich and important experience. As one of the older pilots at the squadron said to me in confidence after the debriefing, it surely made me a better pilot. He concluded perspicuously that one simply have to experience something similar to what I just had experienced before 'knowing what it is all about', one must only hope that it goes as well as it did for me and that one gets this kind of experience before ending up in an even worse situation. My erring judgment had given me some thoroughly comprehensive experience presumably endowing me with a less erring judgment.

Perhaps as a consequence of already being immersed in the academic analytical environment on my way to become a pilot engineer (or simply because of having an analytic personality), the failed looping experience became in addition a trigger for an increased focus on the human component in my upcoming studies of human-systems interaction. My earlier rather narrow focus on MMI and computer displays simply had to be widened to include physiology and psychology, a combined topic that as far as I am concerned is covered by the notions of Human Factors (HF) and Human-Systems Interaction (HSI). Furthermore, I found it very interesting to study what kind of system designs and technological properties that work in the best harmony with the human component. The two spectacular crashes with JAS39 Gripen prototypes that had happened shortly before my little incident were presumably also rather influential on this interest of mine. These crashes that I refer to were the failed landing in 1989 at the Saab airfield in Linköping (a crash nick-named 'the most expensive rotary cultivator ever') and the air display crash in 1993 on the Långholmen island in central Stockholm,[22] both had something to do with the computerized flight-control system required for the aerodynamically unstable Gripen fighter. Part of the explanation of these crashes was that the pilot, an experienced test pilot, did not interact well with the fly-by-wire flight-control system, thereby aggravating problems during the build-up of both crashes. The story was that the pilot had intuitively given control input to the flight-control system according to what probably would have been appropriate if flying a conventional fighter, but in these cases, the input added to the control-burden of the computerized control system making it unable to cope with the dynamics. Some say that the best action in both these cases probably would have been to let go of the stick early in the build-up and rely on the flight-control system to sort out the situation and stabilize the flight path. With my own close encounter freshly in mind I could not avoid sympathizing with the test pilot, I would not have liked the idea of stepping out of the control loop like that during my incident. The intuitive reaction for a pilot during a hazardous event is not to become passive, on the contrary, it is to become increasingly active because in the role of being the captain of the vessel it is your responsibility to manage and regain control of the situation. During stressful emergency-situations, intuitive low-level and hands-on aspects are more important than ever. It is in such situations that the kind of human-systems interaction I focus on becomes tested to its limits, and I cannot resist crediting the Viggen with having been rather intuitive. So, the question became, what are the major differences from a human controller perspective between the modern computerized Gripen and the 'steam-engine' Viggen?

[22]Fortunately, no one died in either of these accidents. The pilot survived the cultivator experience with no more than a broken arm and ejected safely at Långholmen (yes, it was the same pilot). However, a woman was unfortunately burned seriously in Stockholm.

After finishing my Master of Science degree and formally having converted to being a pilot engineer I did, however, not return to the squadron full-time, which in the end made me not become trained as a Gripen pilot.[23] As a freshly graduated pilot engineer I was recruited to a newly established department at the armed forces headquarters to work with electronic warfare and data support systems both for the Gripen and the updated Viggen versions, while continue to fly occasionally. Shortly thereafter, the armed forces reorganized and the air force tactical command was established, an organization that included a development department to which I was recruited. In its different forms and organizational residents, the development department is where I have been working since then, with longer interruptions only for conversion training on the interceptor version of the Viggen (JA37) and for a one-year staff-officer's course at the National Defence College.

3.3 Human factors in aviation, and in car driving

At the Air Force development department, I participated in several military 'meta-scientific' studies. These studies were commonly conducted as joint collaborations between national defense related authorities but also with international authorities as well as with the industry. The label meta-scientific is chosen here because some of these projects, especially studies internal to the armed forces, cannot righteously be maintained as having been scientifically rigorous, which neither was intended. This, from a scientific perspective kind of relaxed approach is required because these studies aim to provide decision support for issues that are either too complex or too hurried to allow for in-depth studies according to rigorous scientific procedures. On the other hand, for me personally they implied studying, and thereby gaining of some kind of deepened insights about, a broad spectrum of interesting issues. The premises were often to provide the best possible insights based on best possible grounds according to available time and resources (i.e., as scientifically rigorous as the conditions allowed), with the underlying assumption that for these particular (military) subjects a more comprehensive understanding did not exist anywhere else, thus contributing valuable and exclusive understanding regardless the rigor. In particular, they were not 'anti-scientific' as they were all conducted with a scientific spirit that naturally included measures to refute the underlying assumption of exclusiveness. What is relevant about this kind of studies in the present context is that from my personal perspective it allowed me, fairly much at my own discretion, to focus on what I found to be most important regarding

[23] I was on the verge to begin conversion training on the Gripen when the overall air force planning process shifted priorities. Shortly after that I got the opportunity to continue with further academic studies, of which this thesis is a concrete result.

the studied matters. Unsurprisingly, I had no problem in relating most of the work to human-systems interaction issues, as the military world almost exclusively is about people and technology. Especially, while conducting these studies I had the great fortune to become part of the human factors in aviation community, initially participating as a subject matter expert and later as an analyst of the human component and of human-systems interaction issues. Overall, I believe, this work gave me a good overview of the field.

While flying stories and aircraft tend to capture the interest of most people, it is personally familiar to fewer than cars and driving. As these two areas both are about human beings controlling vehicles I will therefore continue the present description of experiences from analyzing the human component by following up on the discussion about traffic safety and car-driving that was initiated in the introduction. In order to describe intelligibly experiences of human factors in aviation issues, I will now discuss human factors issues in car driving.

If driving skills and related decisions are divided into three levels, then what I meant earlier when talking about the two-way connection between low-level and high-level decisions was the seemingly often neglected but crucially important impact of low and medium level skills on high-level decisions. By low-level driving skills, I mean those craftsmanship-like, mostly tacit, and highly practice dependent kind of skills that facilitate manual hands-on maneuvering, such as effortlessly being able to shift gears smoothly (with clutch and manual gears, of course), or intuitively keeping the car steady when passing through muddy snow or when being close to aquaplaning. By high-level driving decisions, I mean the mainly analytical and rational decisions that are connected with choosing the right direction (knowing where to go) and the following of traffic regulations. These two levels, the highest and the lowest, correspond rather well, I believe, with the model of the intuitive system 1 and the rational system 2 (Tversky and Kahneman 1974, 1981, Stanovich and West 2000, Kahneman 2003, 2011). While high-level decisions primarily are the result of deliberate and effortful utilization of system 2, these decisions are, probably to a greater extent than we like to admit, inherently depending on experiences and internalized knowledge (i.e., intuitions and feelings, heuristics and biases) associated with system 1. The two-way dependent character, the interconnectedness between these two systems appears to me significant enough to merit a level of its own. Naturally, I call this, the medium level, system 1½, covering the complex borderland between system 1 and system 2 (the model of Tversky and Kahneman, and other models depicting two distinct characteristics of human thinking, are discussed further in ch. 9.1). My classification of a system 1½ does not introduce anything essentially new compared to the more established notions of system 1 and 2, but it covers certain aspects I

consider particularly important for human interaction with dynamic real-world systems such as vehicles.

Essentially, medium-level (system 1½) decisions have, as I see it, a component of rational system 2 aspects, but they are decisions thoroughly affected by the tacit and intuitive system 1 that, almost exclusively, is trained by low-level skills coming from hands-on experiences. Medium level decisions are, for car driving as being the current example, about whether it in a specific situation is possible to increase, or is necessary to decrease, the speed when approaching a bend. They are about where to position the car laterally (i.e., select track) through the bend, or how to behave specifically in a real situation in order to blend smoothly into the traffic rhythm, and so on. These decisions and actions are, I believe, semi-conscious, slightly rational and thoroughly governed by low-level experience and 'gut-feelings'. A decrease in such medium-level skills is presumably significant for the case of the failed looping. The high-level knowing-that I could go through with the looping under the present conditions was all right. The low-level knowing-how to maneuver the aircraft through the looping was in some sense also all right (I had made a few successful aerobatics maneuvers just before the failed looping, but during those maneuvers I was probably focusing more explicitly on what I was doing hence applying more system 2 resources). The problem appears to have been that my intuitive awareness about the actual quality of my lower-level knowing-how (my situation awareness regarding my intuitive skills) was not working in concert with the actual quality of my lower-level know-how, implying that my intuitive system 1 did not alert and call for higher-level resources when it obviously was required. This is what I mean by medium-level, system 1½, decisions. They are actions taken and decisions made as the result of an intuitively triggered ('gut-feeling') use of intuitively grounded (experience based) rational (system 2) resources.

However, low-level skills and the quality of medium-level decisions are perishable, their decay depend on how these skills are maintained, which in turn is affected by the character of the controlled system. With the present technology development trend, a trend that tend to push the control task for human beings to higher levels of abstraction, the ability to make insightful decisions on those higher levels becomes reduced because such decisions are required to be grounded in lower-level experiences not allowed (often not even possible) to be experienced anymore. The most important aspect of safe driving (and of safe flying), I maintain, is the intuitive presence of mind, the situation awareness about, and actual quality of, medium-level decisions. By intuitive presence of mind, I mean primarily an intuitive knowledge of the limits of one's intuitive skills. This is because it is not, more than indirectly, a conscious high-level (system 2) choice to violate a regulatory speed limit on a certain stretch that causes you to run off the road or crash into something. It is rather the semi-intuitive medium-level decision not to adjust the speed

continuously all the way along the stretch in concert with your low-level maneuvering skills and overall situation awareness that finally does it. That is, insufficient maneuvering skills → inappropriate system 1 'gut-feelings' → ignorant system 2 decisions → unmanageable maneuvering tasks (mostly system 1 skills) → uncontrolled skidding, or similar. Hence, inappropriate medium level decisions are the most dangerous aspect of driving, and such decisions are necessary to make continuously even if regulatory speed limits never are broken. This assertion, provocative for some people, requires thereby further scrutiny and a clarifying scenario.

In Sweden, perhaps particularly in the northern regions, regulatory speed limits used to be quite 'generous' (presumably because of the comparatively low traffic intensity that didn't require stronger restrictions), meaning that it was more common than today that environmental conditions (e.g., ice, snow, and wild animals) made it impossible to go as fast as actually was permitted. Of course, this resulted too often in incidents and tragic accidents, but in light of the present discussion there are, regardless, a few interesting aspects about this historic situation. Firstly, it enforced drivers to get to know the limits of safe driving first hand and to maintain some kind of edge awareness. Otherwise, they self-evidently ended up as traffic accident victims. Fortunately, in the northern parts especially, this was mostly a matter solely regarding the victim (disregarding other societal costs such as consumption of hospital resources) because the low traffic intensity made it unlikely for anyone else to become involved. Secondly, it was assumed that drivers were to take appropriate responsibility and not before long after getting a driver's license develop sufficient experience to cope with these conditions (cf. Reed 1996). Today, the apparent attitude of the Swedish traffic safety authority (in the present context perhaps best described as favoring a model-based approach to safety) is to introduce a plethora of speed limits, with much more fine-grained intervals that also often are lower than before. This is done despite the fact that the roads now are wider and that modern cars are overall much better regarding both passive safety (i.e., crash resistance) and active safety (i.e., traction performance). The trust in that drivers aim to maintain a decent level of experience and the idea that they are to take responsibility for their driving has apparently been shifted to a trust in general statistics and to the idea that safety is the responsibility of the traffic model. Irresponsible driving is met by increased regulations.

Do not get me wrong, the problem of drivers not taking required responsibility thereby causing accidents must naturally be addressed, especially since the increased traffic intensity today makes it more probable that also responsible drivers become involved. For this purpose, refining the traffic safety model (e.g., increased regulations) and improving structural safety (e.g., the building of separate carriageways) are important aspects. What I am getting at is that tightening the speed regulations and focusing on conformance to this particular aspect of the traffic safety model might

actually have an unfortunate (as in undesired and very dangerous) effect on what kind of responsibility drivers can and choose to take on.

Essentially, safe car driving is about one thing only. It is about a continuous and skillful control (implying mostly low-level control actions and medium-level decisions) of two tightly coupled aspects: track and speed. These aspects are connected by imperative physical relations, meaning that they are undeniable. At a certain speed, for certain road conditions, the track becomes going straight ahead, no matter how the wheel is turned. Simultaneously, track and speed are aspects that normally are limited individually only by deniable regulatory relations, possible to violate either deliberately or by mistake. For the ability to judge intuitively undeniable implications of control input for one of these two aspects, it is required to know intuitively also the implications of the other. Without such knowledge, a successful drive is the result of using high-level decisions to remain within a set of regulatory limits that happens to be sufficiently restricted to avoid meeting an unfavorable combination of undeniable physical limits, more than the result of using low and medium level decisions responsibly to achieve a safe trip. The problem is that there are combinations of physical limits forming real edges for safe driving that may be reached regardless the regulatory limits, not to mention edges connected to other situational aspects. For the sake of argument, I maintain therefore the perhaps somewhat controversial standpoint that for car drivers the taking of responsibility for vehicle control according to the real physical limits that occur in the interaction between track and speed is of higher priority than the taking of responsibility for model-based aspects such as keeping regulatory speed limits. Reality comes before theory and it becomes dangerous when the model is considered more important than reality or when the model obscures real-world aspects. The current trend implies, unfortunately, both. We have regulatory speed limits that usually are possible to keep at leisure on high-quality road conditions with high-performance cars equipped with traction control systems that in addition obscure the fact that sometimes the performance edge (ch. 9.3) of the car is touched regardless. There is simply no need for careful and insightful track selection, implying that safety-crucial medium level skills are never exercised, a fact that have (at least) two undesired consequences.

First, if there happens to occur conditions that require local adjustments, beyond the regulatory limits (e.g., bad weather and local road conditions), drivers are likely to not realize this because their intuitive systems do not signal warnings calling for rational decisions to adjust behavior according to their low-level skills, and they will probably continue as usual. Drivers are trained (conditioned) to drive mainly by high-level decisions (i.e., to follow the regulatory limits) and never challenge their medium level skills thereby never developing any, which implies an increased (local) accident risk if, for whatever the reason, a real physical performance edge in fact is reached.

From a general perspective, the trend is beneficial (just like the blindfolding of the plateau people, ch. 1.4.3). Despite the fact that traffic intensity has increased significantly, simultaneously there has been a significant decrease in traffic accidents.[24] However, the lack of accidents, or a reduced accident count, is not the same thing as a system becoming safer (e.g., Perrow 1999, Reason 2000). The threshold for accidents is raised, with a lower accident count as the result, but with the consequence that outcomes are more matters of detached probability than of involved responsibility.

The second consequence is therefore that a too strong focus on regulatory aspects has a tendency to shift the sense of responsibility from the involved perspective of the driver to the detached perspective of the safety model. The more specific the regulatory limits, the more likely becomes the interpretation that a too generous regulatory limit was the cause of an accident, when it in fact always is the failure to manage properly the coupled physical aspects of track and speed that make accidents happen, irrespectively of regulatory limits. Furthermore, detached measures such as speed observations are what easily can be collected and analyzed, implying that conformance to model-based limits becomes the official determinant of safety and of a responsible behavior, while in reality it is the other way around. It is possible for a driver to remain well inside all known regulatory limits, and thereby be considered commendably responsible according to model-based values, while being completely unaware of real-world edges thereby having no means for taking real-life responsibility. This is arguably what the contemporary view of traffic safety looks like. The detached traffic model has become more important than involved control of the car and responsibility has become defined by model-based values, both from the perspective of the authorities and apparently also for drivers that seem to assume that the traffic system is safe enough to allow spending valuable attention on their smart-phones.

In addition, when feeling safeguarded by the regulatory system, without low-level 'gut-feelings' of real and inescapable physical edges, the socially made-up regulatory limits may seem arbitrary and lose their sense of urgency. This may, paradoxically, make people more inclined to disregard (i.e., make high-level decisions to violate) the regulatory system they feel safeguarded by. Without ever sensing the real limits, they feel safe anyway. Drivers are handing over the responsibility of their own and fellow road-users safety to the system, hiding behind a model-based predictability that relieves them from the cumbersome work of developing and maintaining low-level skills and edge awareness required for the taking of responsibility.

[24]While writing this I read in our local newspaper that the number of lethal traffic accidents in Sweden has reduced continuously after peaking in the middle of the 1960s. This is obviously for good and it may appear to be the result of regulatory measures and a tightened safety model, but it may also be a marginal effect of technological advances (cf. Footnote 7, p.21)

Contemporary technology development appears, in addition, to follow the model-based approach and focus on safeguarding against consequences of inappropriate low-level skills. The problem with this approach is that such safeguarding tends to imply efforts to remove the need for these skills, a strategy that, according the present discussion, obstructs the gaining of medium-level skills that are required for insightful high-level decisions. The regulatory or technology-achieved reduction of the need for medium-level skills results in reduced maintaining of medium-level skills, which in turn calls for even more regulatory and technological measures to further reduce the need for low and medium-level skills. This is a vicious circle effectively removing the most important aspect for vigilance and ability to discover and recover from real hazards, the medium-level skill to control track and speed according to rich and varied real-world local conditions. The essence of this discussion, and the object of experience aimed to convey, is about the character of the involved perspective of a system controller and the paradox within human rationality. The quality of our rational decisions is thoroughly dependent on intuitive knowledge about system characteristics. In order to be responsible on a rational level we must be intuitively aware of what it is all about, which require personal hands-on involvement.

3.4 Desired end-states in the fog of war

If not before, but certainly during and after the staff-officers military training course, the delicacy of matters and the vast amount of overall social and societal dilemmas associated with military operations and the use of advanced military technological systems became tangible. Whether the effects of using the technology are for better or worse, are in military contexts often impossible to answer unambiguously. Explicit methods to cope with uncertainties of vastly different kinds are therefore required, planning and assessment tools cover aspects that reach from technological nuts-and-bolts issues to human values and international politics. While military activities often are thoroughly dependent on technological systems thus a prospect for systems analyzes according to hard systems thinking and engineering solutions (a basis for the birth of operations analysis, ch. 5.1 and 6.1), the conduct of operations is mostly referred to as an art form, a social endeavor. The bottom line of this passage is that the notion of an art form is beneficial for keeping the involved perspective of human system controllers in view. For arts, as opposed to science, it is allowed to use notions such as 'sensing the solution' or 'having a feeling' that this or that is the right way to do it, and for artists it is in fact expected that they have unique virtues that make them worthy of being artists. Artistic virtues are in addition to a great extent uncontroversially stated as impossible to replace by technological workings. For military operations the artists are the military commanders,

the battlefields their canvases, the military system with its technologies their brushes, and traditionally, gunpowder and bullets their paint. Today, information and other means are becoming more common on their palettes.

Command and control methodologies are usually centered on an explicit process for developing and communicating desired end-states, with related goals (e.g., decisive points) and centers of gravity (at least by military forces influenced by the classic military thinker Carl von Clausewitz). Centers of gravity and decisive points depict overall intentions and significant aspects for reaching goals that are supposed to achieve the desired end state. The notion of 'the fog of war' (also from Clausewitz) is sometimes interpreted as describing the inevitable uncertainties associated with waging war, much as a result of having intelligent adversaries. The traditional military system and its command and control structure has evolved partly from coping with the fog of uncertainties, by providing some kind of flexible freedom of movement for subordinate commanders while still having them working towards goals set out by the commander. Communication of these end-states and centers of gravity from higher command to lower has, however, always been afflicted with a risk of making issues detached and stereotypical, not least because orders are often conveyed as condensed written messages deprived of rich details (traditionally because of a very limited bandwidth – think about mounted messengers carrying sealed letters). As one example of a construct that, arguably, aims to maintain some kind of involved perspective while the written orders transcend through the chain of command is something that usually has the label 'commander's intent'. As a complement to skeleton descriptions of important rich aspects, the commander's intent is added as an explicit effort to convey what the commander mean by his orders, in terms supposed to be comprehensible by fellow commanders. This section in a military order provide means for the subordinate commanders to gain some kind of insight about the involved perspective of the superior commander, means for understanding artistic intentions. The involved perspective is in addition assumed mutual, ensured by the military hierarchy system. It is assumed that a superior commander would not be a superior commander unless sufficient experience exists to ensure sufficient insight about the involved perspective of subordinate commanders. The mutual trust between commanders fostered by the hierarchal system does also counteract orders being solely one-way communications. Formal orders and the chain of command is strictly hierarchal, going from the top to lower levels, to ensure a coordination of efforts and a striving in the same direction. While the actual and detailed execution of the work is based more on a high level of trust in the experience and skills of each subordinate component, to ensure the maintaining of flexibility, robustness, and quality of effects. As a consequence of a fundamental trust in subordinate experiences and expertise, the understanding of the situation possessed by subordinate commanders is of

crucial importance for orders issued, which is a kind of information communication process going up the chain of command, for example in the form of intelligence reports. The key issue here is the mutual trust between commanders, a trust in their respective expertise in controlling their allotted sub-systems, but what about common system operators?

For Clausewitz himself 'the fog of war' may have meant to depict something else than externally implied uncertainties, he may have referred to an internal cognitive fog caused by the life-threatening stress that within combat environments makes the understanding of the situation poor (Lieberman *et al.* 2005).[25] Between commanders and common soldiers, historically the mutual understanding of each other's involved perspectives has perhaps not been that good. The stereotypic military leadership, insensitive, strict, detailed, and to be obeyed without questions, might be a consequence of this cognitive fog. If common soldiers were allowed to think, they would probably refuse to perform tasks likely to get them killed. Hence, the common soldiers were supposed to, and were made to, act like predictable machines. They were expected to accept that the commanders knew what 'had to be done' in order save the virtue of whatever values the war was about. When technological machines became part of the scene, soldiers were set to control these things, while still being expected to, without thinking, do nothing but make the machines perform the tasks as ordered because they were still likely to be killed when enemy opposition aimed to destroy the machines. Seen from the perspective of the commander, the system operator was a component of the machine, not really expected being endowed with artistic virtues. As commanders also were the ones ordering the development of these machines (at least financing it), little effort was, presumably, put on understanding the involved perspective of the operator.

Perhaps there is a pattern to all this, a thoroughly ingrained tradition among military commanders, business managers, and social authorities, to implicitly consider human system operators as machine components? No matter the reason for or the plausibility of this statement, the pattern appears to be repeated on several levels. On the highest level, in the modern society as a whole, people are expected to, and are made to, control technologies and systems according to orders without applying autonomous thinking because otherwise they might end up doing things against the commanders' intentions. The problem is who is the commander? Furthermore, to what ends and according to what values are the war fought? System operation is in fact a kind of war, at least in the sense that we often put our lives at stake when following these orders. Therefore, to know that values are sound

[25]Clausewitz never used the exact phrase 'the fog of war'. Lieberman et al. (2005, p. C7) states the original phrase as “all action takes place, so to speak, in a kind of twilight, which, like fog or moonlight, often tends to make things seem grotesque and larger then they really are”. For me, this original phrasing can mean both external and internal fog.

should be of prime interest, before blindly following orders. However, because of the contemporary view of usefulness, today orders are often to align human behavior with detached and predetermined models. Governed by the paradigm of technical rationality, what is useful is determined by measures defined by stereotypical models (often economic). Effectively, this view promotes detached calculative models to become our commanders and it transforms people into cannon fodder soldiers not expected to think by themselves. The scariest part of this development, as I see it, is not that orders are given by superior commanders, but the fact that today, the supreme commander is not human anymore but merely a calculative model.

My view is that people should be commanders (although everybody cannot be the supreme commander) and technology our soldiers, and as commanders we are supposed to be artists with our own virtues, not predictable automations. In order to establish such a role for human system operators we must, to begin with, allow ourselves to be artists and credit ourselves for our artistic qualities (e.g., for having creativity, for being curious, and for not being predictable). In addition, we must design the systems we are about to control such that they support our artistic virtues instead of suppressing them. The traditionally detached relationship and lack of insight from the perspective of commanders about the involved perspective of technology-controlling soldiers, the perspective that foster the favoring of predictable human behavior is prevailing, arguably. The relationship prevails in the relation between technology constructing designers and system users, between politicians designing social systems and citizens, between managers designing business systems and company clerks, between manufacturers designing production systems and workers, and so on. The detached perspective is perhaps not chosen explicitly and deliberately, but it follows because of an exaggerated focus on predictability and model-based definitions of desires and values. What all these kinds of system designers seem to have in common is nonetheless an implicit strive for predictability, also for the human component.

It appears therefore as there is an urgent need for a shift in viewpoint. However, if technology designers (implying all the kinds of designers mentioned above) are to begin trusting the system operators that are to become promoted to system commanders, if the mutual trust required for subordinate obedience is to occur, designers must stop giving insensitive, strict, and detailed orders to be obeyed without question. Trust and responsibility are two sides of the same coin. If superior commanders want to have their subordinate commanders to fight cleverly and creatively towards desired end-states by taking full responsibility to accomplish given tasks, they must treat the subordinate commanders as autonomous individuals with artistic virtues, not as machine components. The superficially desirable aspect of having people behave predictably is a chimera, a delusion, with undesired consequences that greatly overrules its

benefits in the fog of war. What really is desired is that people act more according to commanders' intentions than according to specific orders. The flip side of completely predictable behavior is stereotypical and completely irresponsible behavior, which obviously is undesired because the slightest change in conditions or a different nuance in meanings will make predicted behavior irrelevant and specific safety precautions brittle. Moreover, to be able to act responsibly among rich real-life conditions it is necessary to have means to build incentives to act responsibly, which require enough experience of acting autonomously to know what things is all about.

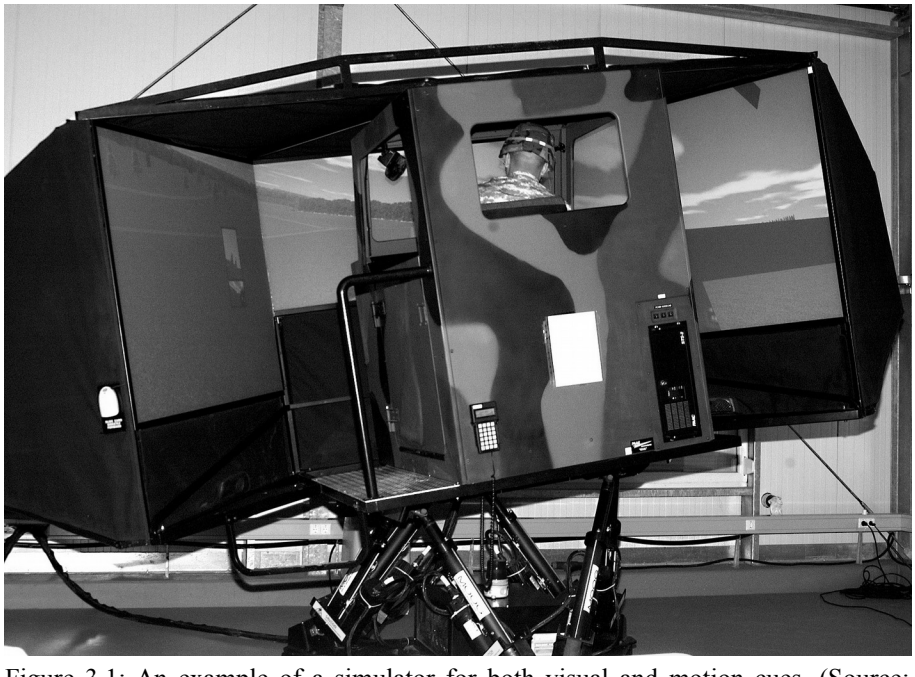


Figure 3.1: An example of a simulator for both visual and motion cues. (Source: wikimedia, public domain)

One way of gaining general experiences is to exercise explicitly likely scenarios under likely conditions. Two problems are then: what scenarios are relevant to exercise explicitly and to what extent are experiences from artificial training situations relevant in real-life situations? It is in this context that modern simulators provide unprecedented means for practicing situations not possible to experience without them. Although, at the same time, simulators constitute unprecedented means for shaping the thinking of system operators and commanders to align with predetermined aspects as well because they may be mistaken for being the real thing. Hence, simulators are great facilitators for skill development, but they are simultaneously double-edged swords requiring thorough scrutiny in order to avoid cutting oneself.

The benefits of simulator training are obvious. With simulators, it is possible to experience situations much too dangerous or too expensive to set up in reality. Such scenarios are today possible to simulate with impressive fidelity thereby giving the impression of achieving relevant first-hand experiences (the simulator in Figure 3.1 provides both a synthetic visual and physical environment). In particular, it is possible for system operators to become familiar with aspects of the controlled system not possible to experience otherwise. On the other hand, simulators are not the reality, no matter their fidelity. They are always models of reality, in the form of modeled system environments and a simplified model of the system of concern, thereby not providing real-life experiences. All models are wrong (ch. 8)! Differences between simulated and real experiences are perhaps particularly significant for complex interaction issues that might occur between the system of concern and rich real-life environments. Error handling procedures for pilots may be used as a straightforward example.

When practicing handling system failures in an aircraft simulator, failures are often handled one by one. For example, loss of air data input. In a simulator, such a system breakdown is straightforward to produce by shutting down the simulated air data provider. The recovery procedure becomes then to find a workaround of the problem, which is to identify what systems that depend on the air data provider, shut them down (or ignore them), and ensure to get necessary information otherwise. A perfectly valid scenario and thereby a relevant experience, if the air data provider breaks down as a solitary event. Emergency checklists often appear be produced according to such simplified scenarios. In reality, however, a plausible cause of loss of air data tends to have richer implications requiring a more comprehensive recovery approach.[26] For example, if the loss of data is

[26]In fact, it is possible to wonder whether simulators are designed according to the checklists, or if checklists are designed according to the simulators, and where reality comes into the picture. Traditionally, emergency checklists were developed from trial and error. Today, when predictability is considered the hallmark of successful engineering and when real-world tests tend to be exchanged for simulator tests, the answer to what is developed from what is not obvious. Consider for example the Mercedes A-class incident from 1997. A Swedish motor-magazine issued what was called an 'elk-test', an evasive maneuver to avoid a moose at highway speed. The newly developed Mercedes A-class flipped over during this test, causing great embarrassment for the renowned car developer. Allegedly, some traditional test-track and road tests had been replaced by simulator trials. The embarrassment comes from the fact that, essentially, a well-built car should not be possible to flip over, unless 'helped' by a slope or a bump in the road. The elk-test showed, however, that a fairly simple real-world maneuver may be richer (include more complex dynamics) than stereotypical simulator scenarios. As some additional spice to the topic, in 2012 the same magazine managed (almost) to overturn a new Jeep Grand Cherokee in a similar setup. One wonders about the ratio of simulated tests for this car. Regardless, the solution provided by the car industry is to equip these cars with more advanced traction control systems. Would it not be better (more resilient) if cars were made physically stable, instead of compensating physical instability with a computerized control system. I mean, cars do not require the physical instability in the way that modern fighter aircraft do. So, why make them physically unstable and rely on detached software models to cope (cf. Airbus and Air France AF447 in ch. 12.3)?

caused by icing conditions choking the pitot-tube, it is likely that other systems are affected as well, including fundamental fight-control system components such as the control surfaces (e.g., ailerons and rudders, not to mention the wings). In the real-world situation, the first indicator of a problematic situation might in fact be a secondary problem, making the trained formal procedure for the identified system failure (and the correct action according to regulations) coming second to responsible and skillful maneuvering to safely get away from the icing conditions. The flip side of simulator training and formal procedures is that, inescapably, it becomes to establish a behavior relevant to predetermined assumptions and model-based values. It is a training that often is better than no training, especially when available timeframes are very short, but if, and only if, there is a beneficial balance with real world training and a sensible scrutiny of what aspects the simulator fails to convey. As for all models, the true value of their explanatory powers depends on how well shortcomings are comprehended. Without such comprehension, powers become exaggerated and the modeled explanations will obscure true aspects. This discussion results in what could be stated as a high-fidelity simulation paradox. When the fidelity of the displays (i.e., the visual, auditive, and, sometimes, tactile displays) on which a simulator is based becomes too good, it becomes more difficult to distinguish the fact that the world conveyed by these displays still is a simplified and stereotypical model of the real world. Trained behaviors that become confirmed as appropriate in simplified simulated environments risk thereby become repeated in real-world situations. This phenomenon is called negative simulator training. For pilots this kind of negative training may for example show as digital handling of the stick (i.e., an on-off kind of maneuvering characteristic) as a result from having no physical forces in the simulator (e.g., no unpleasant G-load), with increased aircraft wear as one undesired consequence. Another example I have heard of is that urban combat training with laser-game equipment may lead to that soldiers take cover behind plaster walls during live firing. Based on the idea that the benefits of models depend on a balanced knowledge about strengths and shortcomings, the computer-game-selling aspect associated with high fidelity simulators may be counterproductive because high fidelity obscures unavoidable differences compared to reality. Low fidelity simulators may actually, in that sense, shoe to be better for training because they make experiences intuitively known to be stereotypical and therefore intuitively known to require explicit adaptation to real situations, thereby resulting in contextually more relevant actions.

The bottom line is that models are always simplifications made according to a specific purpose, and simulators are models. Experiences gained in simulators become thereby shaped according to the purposes that foster the models. Presumably, the purpose fostering models making up simulators is to convey how the simulated system works. However, they convey, arguably,

more how the simulated system is intended to work than how it works in reality. The difference is crucial. Experience gained in simulators is invaluable, if distinguishable for what it is, and devastating, if mistaken for reality. In the fog of war (among the inescapable uncertainties of reality), the desired fellow commander is not one that act strictly according to how things are intended to work, according to practiced routines and by following orders without thinking. No, truly responsible fellow commanders disobey specifics in orders when conditions enforce so, by striving towards the desired end state as conveyed by the rich meaning supplied by the supreme commander's intent. The way automation and computerization of technology is applied today seems, however, to work in the opposite direction.

3.5 Military-technology vs. military technology

This thesis is about human beings and our role in relation to the technology we use, which relates beautifully (as far as I am concerned) to the topic of military-technology, studied at the department of military-technology (obviously) to which I am affiliated professionally during this research. Military-technology (as one word) is in my view about studying the usefulness of technology in military contexts, which should be distinguished from the study of technology used in military contexts (i.e., military technology, in two words). The topic is about the borderland between military operational art and rigorous knowledge about technology. Military operations are ultimately about achieving social (political) effects, by use of military tools of power. Operational art becomes thereby about controlling these tools such that desired effects occur, which necessarily and naturally implies to control these tools also such that undesired effects are avoided. Military tools include almost always advanced technological systems – military superiority by possession of superior technology has in fact traditionally been one of the most powerful driving forces behind technology development – and these technologies are not seldom of a destroying character, a fact that often makes desired social effects indirect and their achievement a complex matter. In essence, conducting military operations implies trying to control a very complex sociotechnical system. The use of military power and its technologies require therefore continuous reassessments of what effects that actually are desired. This requirement makes, for some perhaps somewhat paradoxical, the military community appear more focused on comprehending social aspects and contextual values than the civilian society that to a great extent seem to focus on values described by formal and detached, mostly economic, models. The crucial importance of situated values that depend on local conditions and contextual interpretations is central to the view of usefulness that is presented here.

– Phase one –

4 Outline of phase one results

In order to scrutinize and reflect on the contemporary view of usefulness (the research purpose) and contribute an alternative definition (the research objective), new descriptions are required for better depicting what presently is considered missing. The exploration in phase one, of the concept of usefulness, led essentially to three results. The first result, labeled R1, is the set of models and frameworks presented throughout chapters 5 to 9. This theoretical contribution evolved through scrutiny of concepts central to usefulness and, consequently, concepts that may be used for explaining or (re-)defining usefulness. One essential aspect common to, and implicit in, all these frameworks is the aim to explicate the involved (the situated, the contextual, the generative) perspective. Focus is thereby on the question concerning *how* usefulness may come to be and not only on the detached (the general, the objective, the calculative) perspective more concerned about *what* usefulness is. The second result, R2, presented in chapter 10, consists of two frameworks explaining aspects of situated usefulness, thus directly addressing both the overall research purpose and the main objective. The purpose of phase two is to complement these theoretical frameworks with empirical input. The third result, R3, briefly sketched in chapter 10.3 and elaborated on in chapter 15, is the identification of plausible reasons for the contemporary view of usefulness.

4.1 Models and frameworks, R1

To begin with, the resulting frameworks from phase one are divided into two main categories, serving different purposes. These categories are:

(R1.1) Detached perspective, ontological frameworks
(R1.2) Involved perspective, generative frameworks

From the detached perspective, the frameworks help to describe the *what*, the technology of concern, and *whom* is concerned; the *where* and the *when*, the environment and the context of concern; and the *why*, the purpose of using the technology of concern. Furthermore, these frameworks may also help depict the purpose of the analysis, at least in terms of where and when values appear as well as where and when purposes can be identified. The

involved perspective describes the *how*, how things come to be in a world essentially described by detached aspects.

(R1.1.1) The worldview

The first framework for detached aspects is the sociotechnical outline, the worldview that distinguishes between the technological system (of concern), the physical environment, and the social environment. This worldview does, in addition, discriminate between a local work domain in which explicit system control takes place and a possibly remote domain of effects in which system effects occur, the latter consisting of environments possibly different in character compared to the former, environments connected only by the physical world and the social world.

(R1.1.2) Technological properties, P1-P3

The second framework for detached aspects distinguishes between: (P1) physical properties, the fundamental material properties of a concrete thing; (P2) functional properties, properties of an artifact explicitly designed for a specific purpose; and, (P3) machine properties, dynamic system workings. These properties are outlined along one the dimensions of concrete complexity and abstract subjectivity.

(R1.1.3) Computerized properties, C1-C3

The third framework for detached aspects explicates characteristic properties for computerized systems: (C1) unobservable and discrete dynamics, the result of covert electronic system workings; (C2) complex and dynamic system models, allowing for abstract modes of operation causing system properties to be affected by presuppositions implemented in software and models; and, (C3) seamless and arbitrary systems-integration, the result of computerized networking and communication capabilities facilitating interaction between systems virtually at any level of abstraction.

(R1.1.4) Character of Utility

The fourth framework for detached aspects describes the tension between system properties with a purpose to achieve desired effects and system properties with a purpose to avoid undesired effects.

(R1.1.5) From conditions to effects, in theory

The fifth framework for detached aspects draws a map that explains, in principle, how conditions become effects, outlining the physical domain and

the domain of results that are calculable (as in possible to analyze from a detached perspective) as well as the activity domain and the psychological domain that constitute the contextual. In addition, the framework relates aspects such as appearance, (available) controls, (actual) control, experience, will, and skill, to these domains as well as depicting a principal consequence of automation.

(R1.2.1) Character of uncertainty, U1-U6, U6.1-U6.4

The first framework for involved aspects explicates the major obstacle for actually achieving intended effects, by characterizing uncertainty as caused by; (U1) scope, incomprehensibility due to amount of aspects; (U2) complexity, incomprehensibility due to complex relations between aspects; (U3) dynamic range, calibration difficulties; (U4) dynamic agility, adaptation difficulties; (U5) lack of rigidity, risk for invalid assumptions; and, (U6) distance, lack of coupling between the work domain and the domain of effects. Distances are in turn characterized as: (U6.1) spatial distance, reduced (natural) feedback; (U6.2) temporal distance, feedback lag; (U6.3) contextual distance, detachment, often caused by spatial and temporal distance; and, (U6.4) emotional distance, a common psychological consequence of the other distances.

(R1.2.2) Character of control

The second framework for involved aspects describes the tension between, on the one hand, a detached control of system properties focusing on predictability, the result of a calculative worldview during design (e.g., as a consequence of technical rationality), and on the other hand, an involved control of system properties focusing on situated controllability, the result of a generative worldview during design.

(R1.2.3) From conditions to effects, in practice

The third framework for involved aspects explains how the map from R1.1.5 appears in reality, *how* conditions become effects in practice, by relating several of the concepts present in the other frameworks to each other. The applicability of the calculable (objective) reality in the form of (system) properties and workings provides a contextual (subjective) utility possibly classified as capability or functionality. Such conditions (before the fact) become effects (after the fact) when they are applied in a real situation. The actual application, the usage/control of conditions, create (objective) results labeled effectiveness or safety depending on purpose of usage, and evaluations determine the (subjective) usefulness of these results. Experiences of usefulness influence future interpretations of applicability. Depending on purpose, and whether the character of control is focusing on

predictability or controllability, the resulting usefulness may be characterized as coming from verified effectiveness and safety only or as also having elements of potentiality or resilience.

(R1.2.4) Edge awareness

The fourth framework for involved aspects turns to the perspective of human users whom actually are using the technological systems of concern. It is a framework complementing the well-known concept of situation awareness with the concepts of system awareness and edge awareness. Situation awareness is specified to mean awareness about the environment in which the system of concern is operating, a meaning considered implicit in the established interpretation. System awareness is then specified similarly as situation awareness, to denote explicitly awareness about the technological system of concern, about its properties and workings, its current state, and its potential state. Edge awareness is defined as awareness about the character of the interactions between the technological system of concern and its environment, interactions that tend to have essentially linear characteristics, until some kind of performance edge is reached.

4.2 Explaining usefulness, R2

The second result of phase is made up of two summarizing frameworks directly addressing the research objective, to present an alternative definition of usefulness, a definition focusing on situated aspects.

(R2.1) The conceptual framework for usefulness

The first summarizing framework makes use of the frameworks from R1.1 and R1.2 to put the concept of usefulness in a beneficial frame of reference in which the situated perspective is made explicit.

(R2.2) The character of usefulness

The second summarizing framework provides a rich definition of usefulness, yet readily graspable, a definition based on the tensions within the character of utility (R.1.1.4) and the character of controllability (R1.2.2). Whereas the definition is illustrated as a two-dimensional grid with a cross-hair sight suggesting that usefulness of a certain technology might be characterizable as having a center of gravity somewhere within this grid, the tensions are believed to inhibit such a simplified characterization. The purpose of using tensions is that they suggest that matters cannot be either or, but are always both. Consequently, the character of usefulness is perhaps better described as

a distribution of technological properties relevant for the different aspects indicated by the tension categories (thereby implying a situation and context dependent surface within the grid as opposed to pinpoint characteristics). Moreover, the cross-hair sight is supposed to indicate that while design efforts usually have a specific aim, the resulting 'hit-distribution' (the resulting character of usefulness) may turn out different from intended and depend on situation and context.

4.3 The vicious circle culture, R3

Actually, this third result from phase one can be seen as an explanation for why the contemporary view of usefulness is what it seems to be. Phase one results facilitate deeper insights and a clearer understanding of plausible mechanisms explaining the situation, here sorted into four bullets.

(R3.1) Reduction of concept richness

The first aspect is that it seems there is an ongoing reduction of concept richness for concepts originally used strictly for human qualities, showing as an established practice of using such concepts to denote technological properties. Chapter 15.1 is dedicated to this matter. The problem is in short that by using these concepts for technological properties they slowly begin to mean nothing more than what technology can be, and the richer human aspects originally referred to loses their relevance. One consequence of this loss is that it facilitates R3.2.

(R3.2) Exaggerated faith in predictability and objectivity

The second aspect is that without means to understand fully the rich and involved meanings associated with human aspects of life, calculative and detached system models of the world become the norm. The concepts we are using to understand human aspects become associated with technological properties appropriately described by calculative and detached system models. When the maps, the calculative models, are mistaken for reality in this way, perfect objectivity becomes readily attainable and predictable outcomes straightforward. The driving force behind this mistake is, presumably, the natural human desire for comprehensibility, a comprehensibility that follows straightforwardly from objective and predictable system workings. One consequence of this exaggerated faith is, however, that it also drives R3.3.

(R3.3) Exaggerated strive to accomplish predictability

The third aspect is that if the map actually is the world, perfectly objective analyses of future outcomes are possible and technology can actually be designed to provide perfect effectiveness as well as maintain absolute safety. The problem is that the map is not the real world! In fact, reality is inhabited by human beings that are unpredictable much because they act according to subjective interpretations of meanings and values. What will be considered perfectly effective and absolutely safe in a local and specific future situation is thereby thoroughly unpredictable. Moreover, the physical reality seems complex enough to be thoroughly unpredictable as well, implying that the world is unpredictable even without including social uncertainties. Regardless the uncertainty, R3.2, facilitated by R3.1, makes us conclude that the natural (and necessary!) human variability is the culprit when things turn out different from what is expected (predicted), and efforts to ensure predictability are therefore increased further, commonly by technological means. The driving force behind this approach is, presumably, the natural human desire for predictability, assumed facilitating comprehensibility as well as both effective and safe systems. However, this attitude makes R3.2 effectively become a self-fulfilling prophecy. Arguably, discovered anomalies should be interpreted as model errors that in turn should reduce our reliance on the map and make us choose to adjust it. Instead, we consider what causes the model to be wrong the culprit (because it ruins predictability, presumably) and try to make reality fit the map. The better we succeed in making reality fit the map, the more the faith in predictability will seem appropriate, an approach also leading to R3.4.

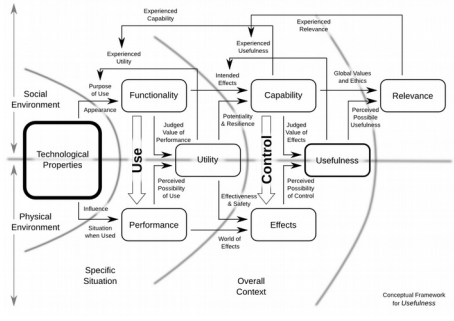
(R3.4) Escape from responsibility

The fourth aspect is that, presumably, yet another desire exists, driving the vicious circle that results from R3.2 and R3.3 mutually supporting each other, namely, the desire to escape from responsibility. The assumed attainability of objectively correct decisions, something that presupposes a perfectly valid map, removes in practice the need for taking responsibility because an indisputable truth is, by definition, correct. Responsibility is required only because there is uncertainty and subjectivity (and conflicting interests), and without subjective decisions there is no need for taking responsible decisions. The delusion of objectivity becomes then a smoke screen possible to hide behind, as responsibility has become someone else's problem. The objective truth is to blame for undesired effects, a truth often embodied as an impeccable system (e.g., a technological, legal, or organizational system).

(R3) The vicious circle culture

In conjunction, these four issues form a vicious circle culture. When the interaction between R3.2 and R3.3 becomes an established pattern and R3.4 has become the (implicit) goal, it is possible to call it a culture, a culture that also seeds a continued reduction of richness for additional concepts related to human beings (R3.1). This calculative culture has the described effect on concepts for human qualities because subjectivity and responsibility are core aspects of being human. Without an acceptance of subjectivity and recognition of the importance of responsibility that follows from living in an unpredictable subjective world, other human aspects become also reduced in richness and value. Intelligence becomes reduced to mean the ability to calculate the so-called objective truth. Consciousness becomes reduced to mean the ability to read the map and keep formal data in memory. Hunches, intuition, emotions, and feelings, become problematic disturbances to, and problematic concepts in, an otherwise formally describable and well-ordered reality. Values become derivable from system properties and model definitions, and so forth. Ultimately, autonomy becomes reduced to mean the capability to function according to predetermined rules derived from predetermined goals in a perfectly predictable world. Arguably, the vicious circle culture is missing a number of essential points.

5 Exploring usefulness



The cluttered small image above is an icon of the conceptual framework model presented in chapter 10.1 (Figure 10.1). Being an icon, it is not intended to be readily readable. It serves instead the purpose of a general map, used in subsequent chapters to indicate what part of this overall framework the discussed aspects or concepts are focusing on.

To find out what aspects are relevant for 'true' situated usefulness of an advanced technological system is not a simple matter. First, to agree on what is true has never been a trivial feat. On the contrary, the nature of truth is an intriguing and constant subject for philosophical discourse. For socially defined matters such as usefulness, in practice, whether a system is truly useful or objectionably repugnant comes down to values and moral standpoints (Rachels 2009). Second, what is implied by situated, situated where, when, and in relation to what? Furthermore, even before starting to explore usefulness for something as concrete as a technological system, the presumably trivial issue of agreeing on what actually is the technological system of concern might be enough problematic to cause disagreement. The trouble with knowing what system to be concerned with is much because some system boundaries are quite difficult to discern. Many technological information-handling systems (i.e., computers and networks) are, for example, troublesome to distinguish from the social system in which they are operated (e.g., Checkland and Scholes 1999, Orlikowski and Iacono 2001). Even while leaving out the question of true values (i.e., true or illusory motives) and assuming that the layout and boundary of the technology in question is straightforward, a clear and unambiguous meaning of usefulness as a concept appears still missing.

Consider for example the following encyclopedia definition of usefulness, "the quality of having utility and especially practical worth applicability."[27] This definition intertwines usefulness with other concepts, apparently equally complex, such as utility and applicability, and it grounds usefulness in the difficult area of values (practical worth). Utility is accordingly defined as "fitness for some purpose or worth to some end" or as "something useful or designed for use".[28] From these definitions, we find that usefulness is to have utility with practical worth and that having utility is to be useful by virtue of properties fit for a purpose. Clearly, these definitions form a circular dependency and neither have a straightforward and unambiguous meaning. The common ground for usefulness and utility appears, however, to be value, or more specifically, having properties with practical worth applicability. Purposes for which the technology of concern may be applied, to which values are connected, will therefore be explored as the first step after defining with what technology we are concerned.

This chapter deals with identification and description of what system we are talking about (the system of concern) and what system (or systems) that constitutes its environment. The chapter addresses what properties these different systems have and for whom the technological properties are supposed to have practical worth applicability. An ontological outline of entities is provided, a worldview that facilitates the subsequent exploration of situated usefulness. Next chapter (ch. 6) explores under the headline of desired effects what practical worth possibly might imply, followed by three chapters (ch. 7 - 9) about how potentially useful properties may come to be truly useful in real-life situations, these latter three chapters are about applicability and controllability. Then, in chapter 10, an explanation and redefinition of usefulness is provided. For the following outline of different aspects relevant for exploring the concept of usefulness, the selected theoretical framing is systems theory and systems thinking.

5.1 Systems thinking

The beginning of modern systems thinking practice is often attributed to cybernetics (Wiener 1948). Particularly, perhaps, it is attributed to the outline of general systems theory (GST) by which Ludwig von Bertalanffy (1950, p. 139, emphasis added) aimed to extend "*exact science*, meaning a hypothetico-deductive system" that, so far, had been almost identical with theoretical physics, to also apply to biology, psychology, sociology, etc. As its name suggests, GST mounted to be applicable to all sciences with promising prospects, it was seen as the skeleton of science (Boulding 1956),

[27] http://www.merriam-webster.com/dictionary/usefulness
[28] http://www.merriam-webster.com/dictionary/utility, both quotes

although the approach came with certain difficulties. Boulding (1956) suggested, for example, not less than nine theoretical system levels in order to cope with certain matters of complexity (Weaver 1948, Simon 1962). The higher of these nine levels align rather well with the four primary levels of emergence (ch. 2.1.3) suggested by Emmeche et al. (1997), mainly regarding matters related to biology and human sciences. In fact, and perhaps not that surprising, classical systems thinking as promoted by theoretical physics is particularly powerful when dealing with concrete physical systems and problems related to analysis and construction of artifacts. The success of the Systems Engineering (SE) movement (e.g., Jenkins 1969) is, presumably, a result of the relatively straightforward applicability of systems thinking powers on concrete physical matters.

Another profession that adopted the tenets of systems thinking was Operations Research (OR), initially dealing with military matters such as optimizing ongoing operations by adding scientific assessments and evaluations to operative and strategic decisions.[29] OR became later more and more associated with management issues and business problems (Churchman 1970, Ackoff 1979a, 1979b). Management and OR are more socially oriented than artifact construction and the social shortcomings of classic systems thinking became evident rather quickly. Systems thinking is, however, an approach powerful enough to function as an ideology (Lilienfeld 1975) thereby potentially enforcing its application too widely, which lead to conflicts within the OR community (Checkland 1983). The controversies regarded mainly the inapplicability of classic systems thinking on social systems, with are systems with a counterintuitive behavior (Forrester 1971) compared to the deducible workings of physical systems and machines. By some, the engineering approach to social systems was considered maladaptive (e.g., Hoos 1976). Continued misapplications of systems thinking on social matters led eventually to the development of soft systems thinking (e.g., Checkland 1985), to critical systems thinking (Flood 1990, Jackson 1991), as well as to the development of corresponding methodologies such soft system methodology (Checkland and Scholes 1999) and critically systemic discourse (Ulrich 1987, 2003).

The concept of a system may be problematic in itself (e.g., Marchal 1975). One example of a definition of the concept is from systems engineering: “a combination of interacting elements organized to achieve one or more stated purposes” (INCOSE 2006, p. 1.5), which is a definition that views the systems analyst as a passive observer (i.e., a detached perspective). The soft systems approach denotes the classic engineering view as hard systems thinking and considers instead systems as purposeful activity systems that explicitly include the systems analyst (Checkland and Scholes 1999). Hard

[29]OR originates from the late 30's when (natural) scientists worked with officers of the Royal Air Force in the UK to create and measure effectiveness of radar-based air defences, a work that A. P. Rowe called “operational research” (Checkland 1983, p. 664).

systems thinking tend to produce statically defined and deterministic system models, while soft systems thinking lead to dynamic and learning systems. Stafford Beer's 'viable system model' (1984, p. 8) suggests that some system definitions are more natural and self-sustaining than others, a model that also includes the central principle of recursion: “every viable system contains and is contained in a viable system”. The notion of system of systems (e.g., Jackson and Keys 1984) is essentially the same concept as the principle of recursion, the view that a system may be seen as an element of its own right in a system consisting of many other elements, as well as being an element having systems as its parts.

Hard and soft systems are rather intuitively connected to (viewed as taking stance in) physical and social systems respectively, and the controversies (within for example OR) can therefore be summarized as a debate about which approach covers the other, a debate quite in line with (and presumably part of) the debate between positivism and interpretivism briefly recapitulated in chapter 2.1. Hard systems thinking is an approach that facilitates a detached perspective, giving an impression of objectivity that appeals to positivists whom thereby use the approach also to analyze social matters. Soft systems thinking enforces an incorporation of subjectivity and purposes (of both analysts and participating individuals) that appeal to interpretivists whom thereby use the approach also for analyzes of physical matters. Presumably, both approaches are inadequate if considered the only one, and the ultimate kind of systems thinking ought to be a combination of both. For systems thinking explicitly, the efforts to promote critical systems thinking and critically systemic discourse may therefore be considered parallel to the ambition of critical realism for philosophy of science, the ambition to join the competing views.

It seems that social and physical systems are fundamentally different in character yet inescapably interdependent thereby forming sociotechnical systems (STS), a term that stem from the seminal work by Trist, Bamforth, and Emery (the process leading to the discovery of the concept is recapitulated in: Trist and Bamforth 1951, Trist 1981). Having technology that may be described adequately with a hard systems approach, within a social setting requiring a soft systems eye, do complicate things, and the standard approach has long been “industrial age thinking”, favoring formal rationality and focusing on efficiency, predictability, quantification, and control (Walker *et al.* 2008, pp. 481–482). It seems at least quite difficult to create relevant descriptions of sociotechnical systems only with notions from either of the two major systems thinking traditions. Marchal (1975, p. 452) provided three criteria of adequacy for a satisfactory explication of the use of system (of concern). First, the basis for identifying and distinguishing between different systems must be clear. Second, the basis for identifying and distinguishing between different kinds of systems must be clear. Third, a

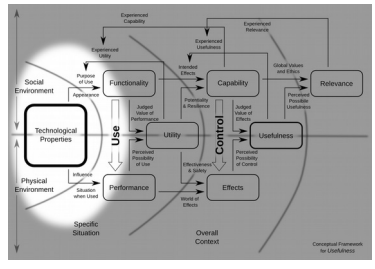
satisfactory explication of a system will not result in everything being a system. The last criterion implies that there are no closed systems.

Implications of sociotechnical systems-theory reach far beyond the scope of the present research, with design principles and methodologies that include business and management as well as politics and ethics (Cherns 1976, 1987, Pava 1986, Clegg 2000). The present focus is, in some sense, limited to specific issue of human control of technological systems, focusing on the 'primary work system' and virtually ignoring the whole 'organization system' and the 'macro-social system', three levels of analysis defined by Fox (1995, p. 95). However, for many technological systems, and perhaps particularly for military systems, such specific control includes necessarily overarching social aspects, for example when analyzing command and control systems (Walker *et al.* 2008). The main reason for the sociotechnical setting in this context is that the purpose of system control, the norms and values that guide and govern incentives for controlling even the smallest gadgets, are socially dependent and evaluated against social norms. For the present purpose the term sociotechnical and its traditional business and industrial management connotation is, however, not optimal because it does not, in my opinion, acknowledge appropriately the dynamic interaction with changing physical environments that is characteristic for vehicle systems. Yet are vehicles and their control undoubtedly dependent on social aspects. Using the hard-soft distinction and categorizing technology plus the physical environment as hard (belonging to the natural sciences) and the social environment as soft (studied by the social sciences), an alternative term would perhaps be socio-natural or socio-physical systems theory.

When dealing with sociotechnical systems (or socio-natural, or socio-physical) a dual focus is necessary, it is required to consider two open and interdependent systems (Fox 1995). The frameworks presented below acknowledge this standpoint and takes the requirement of diversity further. It may be required to consider more than two open and interdependent systems. Inspired by Marchal's three criteria for satisfactory explication of systems, a worldview is presented that distinguishes between the social system, the physical environment system and the technological system of concern, a distinction based on the idea that they are systems fundamentally different in character, they are different kinds of systems. In addition, the worldview distinguishes between different social and physical environment systems depending on purpose, the work domain consists of the sections of these systems that primarily are relevant for control and the domain of effects consists of those that primarily are relevant for system effects, the domains may be considered essentially different systems. Regardless the distinction between these different systems of different characters, they are still connected by the social and physical world systems, included as an explicit remainder of the inescapable openness of the described systems of concern. The explored phenomenon of situated usefulness may then be

described as an emergent property of the sociotechnical system of concern as a whole, a system explicitly distinguished from the entire world system.

5.2 Sociotechnical worldview



To begin with, usefulness is clearly a socially (culturally) dependent property as it is a concept that includes values. Therefore, a worldview suitable for defining aspects of usefulness must necessarily include social aspects. While the social world may be described as a social system, the concept system might be misleading because it tends to give the impression of deterministic properties, especially if classic (hard) systems thinking is assumed implicitly or subconsciously. The wicked character (Churchman 1967a) and counterintuitive nature (Forrester 1971) of social systems may therefore call for another notion than system in order to avoid invalid predictions, leading to disrupting appearances like black swans when that is assumed impossible (Taleb 2010). The social environment could perhaps instead be called field or domain to stress the holistic character of social aspects. However, to function as an interface to benign applications of classic systems thinking it appears required to view the social world as a system, but nonetheless a system with rather special properties and vague boundaries.

Technological systems, on the other hand, are concrete artifacts created by human beings for a purpose and for technological gadgets the boundaries are most often unambiguous, for example, a lap-top computer, an industrial robot, or the last generation military fighter aircraft. For transportable things, the distinction between the technological system and its environment seems natural because the artifact may be moved into different environmental settings. Although, issues with continuously changing and thereby somewhat unknown environments are perhaps especially relevant for vehicles as one of the main purposes with vehicle systems is to have the capability to move around without unnecessary built-in limitations for where to go. However, as technological gadgets are physical things they cannot be analyzed properly in isolation because the physical technology-system interacts continuously with the physical environment-system. Commonly, these interactions are in addition what facilitate core system functionalities. The industrial robot

requires, for instance, to be thoroughly secured on a stable foundation while applying its servomotor powers and vehicles require traction reaction forces.

Other system boundaries may be more difficult to distinguish, for example a portable GPS-navigator system. Is it merely the gadget in your hand, or does 'the system' also include the satellites, the map databases, the map-data collection systems, the data preprocessing systems, and so on, and what about command and control (C^2) systems? Are they merely the technology used to control procedures (i.e., the communication devices), or does 'the C^2 system' include the procedures as well, the rules governing the procedures, doctrines and moral standpoints governing the rules, and so on?

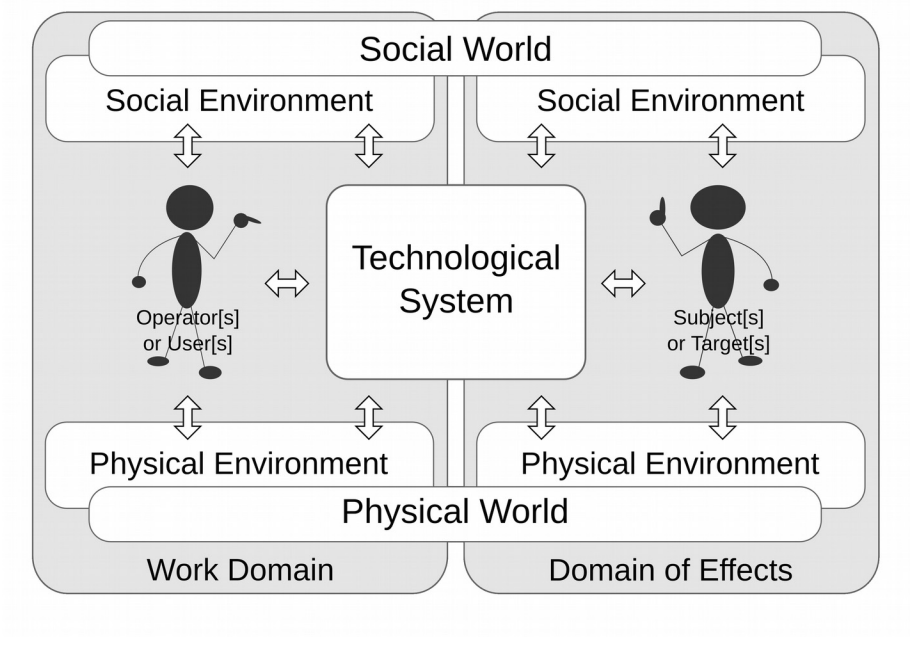


Figure 5.1: sociotechnical system outline (R1.1.1), the worldview

There are, in addition, principally two categories of human beings directly related to the technology of concern. First, there are individuals whom are users or operators working the technological systems. Second, there are individuals whom are subjects or targets of system effects. These individuals may all be viewed as sociotechnical system entities, where some aspects may be describable by hard (i.e., detached) system notions (e.g., physiology) while other aspects might require a more involved perspective (e.g., psychology and sociology). These different categories of individuals have minds and bodies possibly embedded in distinctly different social and physical environments only connected loosely on a global level and by the technological system of concern. Hence, the demarcation of different system boundaries is not at all trivial.

For the present purpose, the world is divided in four principally different categories of systems: the technological system of concern, individual human being systems (people), physical environments, and social environments. In addition, the different categories of people possibly connected with distinctly different physical and social worlds provide the basis for distinguishing between a work domain and the domain of effects, still connected by the virtually boundless physical and social worlds. The layout of the sociotechnical worldview (R1.1.1) is illustrated by Figure 5.1.

The technological system (of concern) is depicted as an ontologically distinct entity, simply because it is the main object of concern when scrutinizing technological usefulness, it is the actual thing to be designed (or evaluated) and used, and it is accordingly depicted in the center of the model. Every non-trivial technology is in addition a system of systems made up of several physical objects with generally well-defined properties thus with comparatively great success analyzable with hard systems thinking. The properties of the technological system of concern is at the core of the present research and these properties are therefore further scrutinized in chapter 5.3 – Character of technological systems, and chapter 5.4 – Character of computerized systems.

The physical environment provides much of the working conditions for the technological system of concern, especially from a concrete mechanical point of view. Environment conditions may for example be weather for outdoor technologies, surface structures for ground-vehicle systems, energy supplies for active systems, and interacting (interfering) objects such as other vehicles and road obstacles for moving systems. Because many properties of technological systems depend thoroughly on such working conditions, design limitations governed by the purpose of the design put constraints on allowed environments for system operation. Such constraints can for example be a recommended temperature-span, allowed vibration-levels, limits for atmospheric pressures, and such like, limits within which the technology is supposed to work according to its specifications. That is, if these specified limits are violated, unknown and probably undesired effects may occur (of which the perhaps most probable effect is the ending of manufacturer guarantee commitments). Interaction with the physical environment is often rather direct, making effects immediate and a lack of necessary conditions directly block further operation. The physical world is inescapable, necessary, and most often rather uncontroversial, but still, perhaps, slightly unknown. Nevertheless, the physical world is in most aspects restrained by factual (and in practice unquestionable) moderating forces. Thus, given certain conditions such as a graspable scope and a reasonable timeframe, the physical world tends to be rather predictable. The physical world can often be described appropriately as 'Mediocristan' (Taleb 2010), a varied but self-regulating world compatible with the Gaussian probability distribution.

The social environment provides most of the contextual conditions for use of the technological system of concern. For example, cultural and socially dependent psychological aspects, norms, values, expectancies, strategies, and such like. Motives for using the technology in question, incentives for trying to control it, criteria for assessments and evaluation of effects, and judgments of relevance, and so forth, are all aspects determined by (emergent in) the social environment. Interactions with the social environment are often more indirect compared to interactions with the physical world and system operation is seldom fully blocked by social issues. The social environment is more open for local interpretations, its aspects are seldom unquestionable facts but instead thoroughly disputable, and may in addition be controversial and even conflicting. Social factors are in fact often possible to disregard, which implies that social rules can be violated, deliberately or by mistake. The social environment is, however, equally inescapable and necessary because without a socially grounded incentive to apply the properties of a system no deliberate system operation will take place. Furthermore, when some people use certain systems (and voice their approval of system effects) others tend to follow, which is one reason why aspects of the social environment tend to be self-generating, exponential, or unbounded. The social world is thereby most often appropriately described as 'Extremistan', the breeding grounds for black swans (Taleb 2010), and in practice, it seems quite unpredictable.

Operator[s] or user[s] are those human beings that are involved in controlling the technological system of concern. This control activity can on the one hand be nothing else than a one-time effort, for instance when fixating a technological artifact as a component in a larger construction, or when setting off a fully automated system. On the other hand, control activity can also be a continuous and perhaps highly laborious manual control work, such as when skillfully using a tool to craft something, or when driving a vehicle in a demanding environment.

Subject[s] or target[s] are those human beings that are affected by (or, more specifically, that are intended to be affected by) effects from operating the technology. For some kinds of systems, users can also be subjects, typical standalone personal computer systems, for example. Other systems are better described as having distinct operators and targets. Sometimes subjects (or targets) can be users for short periods as with ATM-machines that are operated by the banks but under highly regulated conditions used by their targeted customers. That is, operators (the left-hand side of Figure 5.1) are considered active and authoritative system-controlling individuals with access to all means of control provided by the system. While subjects (the right-hand side) are viewed as more passive and controlled individuals affected by the operated technological system occasionally granted access to explicitly designated means of system control.

From the perspective of a human system operator the work domain is everything relevant for operating the technological system at hand, which may depend fundamentally on the kind of technology used. On the one hand, there are computers with primarily stand-alone workings such as word processors. These are typical coherence-driven driven work-domains (Vicente 1990), where external conditions mostly are static requirements. When typing in a local document on your personal computer important aspects are primarily a matter between you and the computer, as long as certain environmental conditions are fulfilled, such as access to power and a sufficiently comfortable place to be. The targeted effect from using the system is a document, ultimately defined by the system representation. Any discrepancies between the computer representation and the user view of things (e.g., Norman 2002) will of course lead to misdirected control efforts (e.g., manipulating appearance directly when changing a format temple would be more appropriate, thereby failing to achieve the truly desired effect). However, the actual result is nothing worse than an imperfectly structured document, and it is imperfect only according to the user who simply has failed to maneuver the technological system in precisely the desired direction.

On the other hand, there are computerized production plants or vehicles. They are examples of correspondence-driven work domains (Vicente 1990), where the continuous dynamics of the system-external world is essential for system operation. When flying an aircraft with a computerized autopilot or flight-control system, it is obviously of utmost importance that the system representation of aspects of the environment and how the human pilot perceive things agree reasonably well. The system operator (the pilot) and the system must, so to speak, talk the same language and understand each other (i.e., the pilot must understand, the system must be understandable). Discrepancies between computer representations and operator views will for correspondence-driven systems just as for coherence-driven systems lead to misdirected control efforts. The crucial difference is that for the vehicle kind of system the targeted effects have a much more physical footprint. Misdirected control efforts have for vehicle systems tangible consequences, such as crashing airliners because pilots fail to maneuver them in precisely the desired direction.[30]

Ecological interface design (EID) is an approach explicitly focusing on correspondence-driven work domains and their specific requirements, which are requirements that come from the immediate couplings between the technology and the system-external world (e.g., Vicente and Rasmussen 1992, Vicente 2002). The term ecological is taking stance in the insight that human beings seem better at understanding natural phenomena than artificial ones, and that we therefore would be better at controlling systems providing

[30]See SAS, SK751, in ch. 12.1, and Air France AF447, in ch. 12.3

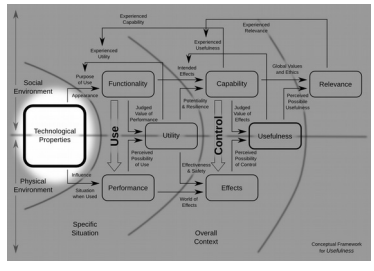
natural feedback that we can interpret intuitively, compared to systems presenting abstracted feedback requiring interpretation with considerable cognitive effort. However, for systems where the technology is used to explicitly facilitate being in environments unsuitable for human beings, the ecological approach becomes counterproductive. For example, computerized flight-control systems are necessary for controlling unstable airframes at high speeds, simply because the physical dynamics are much too quick for a human being to handle. If such a system would provide natural (ecological) feedback, it would obviously be a too rapidly changing kind of information for a human operator to assimilate.

There may also be one additional and further complicating issue, besides the requirement for compliance with the natural world. With some kinds of systems, such as networked or otherwise distributed technologies, it is not necessary that effects from using the system actually affects the local work-domain in any other way than through explicitly designed feedback facilitated by the technology in question, or through complementary technological systems (e.g., sensors and command and control systems). An extreme example is teleoperated (remotely operated) weapon systems, such as armed unmanned aerial vehicles (UAVs) often referred to as drones. For drones, both the physical and social environments surrounding the far side of the technological system (i.e., the remote vehicle with its armament) are significantly separated from the physical and social environments surrounding the operator and the near side of the technology (i.e., the local control station). The everyday example is networked computer applications (e.g., Internet-based client-server systems), where desired effects primarily appear on the client side while authoritative operation mainly is done on the server side, and these sides may be located independent of each other, virtually anywhere on the planet. For sociotechnical systems where sought-after effects appear in distinctly different social situations and physical environments from that of system operation, it is required to complement the work-domain with the concept of a domain of effects. This means that according to the present worldview, the work domain is the near part of the operated technological system together with the physical and social environments where the operation is performed, while the domain of effects is the far part of the operated technological system together with the physical and social environments where the primary sought-after effects from operating the system appear (Figure 5.1).

The different domains may of course be connected through higher order relations, which perhaps is especially obvious for near and far environments of the same kind. For example, the complex intertwining of interests and different stakeholders (i.e., both those benefiting from the effects and those suffering) connects the two different social environments by virtue of the social world, and the global environmental footprint make the two physical environments be connected by the physical world. For coherence-driven

work domains, this explication of a domain of effects becomes, however, irrelevant because the operator[s] are simultaneously subject[s] and the social and physical environments are those surrounding the operator[s].

5.3 Character of technological systems



Many technological systems are complex constructions. Usually they consist of multiple parts of which many in turn are more or less advanced technological systems. Most modern technological systems are best described as integrated system of systems, which are systems that quickly become complex enough to be practically incomprehensible. Regardless the complexity, an analytical decomposition into sub-systems will eventually end with simple parts that principally are nothing but chunks of matter thereby with quite well known properties. These chunks of matter, the smallest building blocks of all technological systems (at least in a mechanical sense, at an every-day level of physics abstraction) have well-known properties because they are appropriately described by since long scientifically established physical laws (e.g., Newtonian mechanics), and thereby they are considered essentially objective. This situation may give the impression that systems consisting solely of such objective physical entities are equally objective and well known, as a direct consequence of combining recursively well-known properties and laws. While the analysis of multiple interacting properties and intertwined laws may be ambiguous by itself, there is more that complicate matters further.

When a technological system is designed and built, the recursiveness of systems thinking analysis is reversed. The analytical decomposition into sub-parts becomes instead an engineering composition of parts into systems, which adds an abstract dimension of subjective purposes as a complement to the concrete dimension of physical complexity (Figure 5.2). Subjective purposes begin already at the material level. Chunks of matter are explicitly crafted into artifacts from materials carefully selected for their suitable properties and deliberately shaped to fulfill one or more functions. This means that besides the set of practically objective physical properties (P1), artifacts have also a complementary set of subjective functional properties

(P2) that depend on the purpose of design. These functional properties are naturally based on and limited by the physical properties of its materials, but they are not completely determined by them. For example, a steal beam is carefully designed to withstand a certain load per length unit (in relation to its shape that in turn determines its flexibility, mounting constraints, etc.). The sought after and explicitly designed functional property of the beam is its weight-carrying capacity, but a piece of it can always be used as an anchor, or a projectile, or anything else conceivable as benefiting from its physical properties.

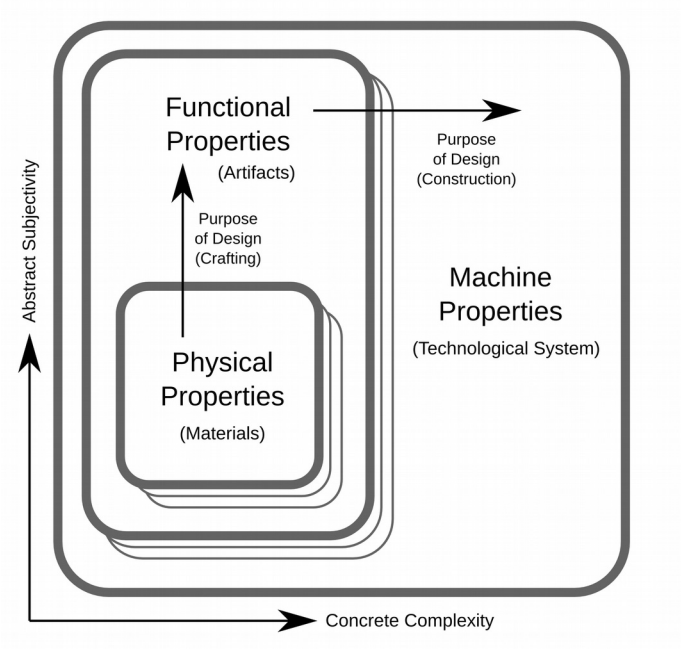


Figure 5.2: The recursive nature of technological properties (R1.1.2)

Furthermore, several artifacts with their respective functional properties can be combined into systems with certain dynamic workings or machine properties (P3). For example, artifacts such as cylinders, pistons, connecting rods, and crankshafts, are combined with, and complemented by, certain chemical and thermodynamic processes to make up a working piston engine. Its sub-part functional properties then become of lesser importance, as long as the system is used within its designed limits and the respective functional parts do their jobs. Only when the engine starts to malfunction, or when it is used under extreme conditions (likely causing malfunction), knowledge of sub-part properties might become valuable. The engineering recursiveness can also be continued, which for a piston-engine system often results in being used as a sub-system part of a vehicle system. As such the engine

constitute but a few of the functional properties defining the limits for the machine properties of the vehicle system as a whole.

Often technological systems are explicitly designed to achieve certain top-level machine properties and many physical and functional properties of sub-systems and sub-system parts become in the process mainly design obstacles or unfortunate consequences. However, when the system actually is used, all properties of all system parts are always present, and they will play their part. Furthermore, in a certain situation anyone of these properties may be what actually matters.[31] While physical properties of 'simple' artifacts often have clear and well-defined implications that in practice can be viewed as objective and indisputable facts, functional and machine properties are defined by subjective purposes of use probably influenced by subjective purposes of design. This leaves open the possibility for novel subjective interpretations, unknown purposes of use, and unpredicted functionalities, perhaps unique for a specific context. That is, functional and machine properties of a technological system are not unambiguously determined by its physical properties, only physical aspect of possible functionality and machine workings are what cannot exceed the limits posed by its physical properties. The three kinds of properties in this model (R1.1.2) are:

(P1) *Physical properties* (materials): e.g., size, shape, weight, strength, etc. These properties are concrete, unavoidable, principally indisputable, rather well known, and may be both desirable and facilitating as well as obstacles for design. They are always present, regardless of point of view and thereby virtually objective.

(P2) *Functional properties* (artifacts): properties explicitly crafted for a purpose, e.g., lift capacity for a steal beam, flexibility of a spring, etc. They are to some extent subjective properties open for different interpretations, which means that an object may besides its desired and purposefully designed functional properties have both unknown (e.g., lethal in the hand of a maniac) and undesired functional properties (e.g., wear, noise, etc.).

(P3) *Machine properties* (the technological system [of systems]): properties that emerge when physical and functional properties interact over some time, they are the dynamic workings that often are suspect to certain working limits (e.g., crankshaft rpm-limits for piston engines, calculating capability of computers, etc.). The recursiveness is evident, machine properties of one gadget-system become functional properties when used as a sub-system in a more complex machine.

[31]Supposedly you buy a laptop mainly for its fascinating performance provided by its machine properties, but you can also always utilize its shape and weight to function as a door-stop, a fact that under the most extreme conditions actually might be what saves your life...

5.4 Character of computerized systems

Computerized systems are principally not different from other technological systems, but they have a few significant characteristics considered highly relevant in this context. These characteristics are properties specific for computers, with specific implications for the interaction between people and computerized systems.

To begin with, for human beings computerized technological systems have completely unobservable and discrete dynamics (C1), unless there is some kind of explicitly designed feedback. The dynamics are unobservable much because functional and machine properties are determined by electrodynamics for which human beings have no useful senses, dynamics covertly hidden within inanimate hardware. In addition, machine properties of computers are largely determined by software, of virtually unlimited complexity. While normal electronics (i.e., not meriting the label computer) naturally can be painfully complex as well, such traditional designs are at least in some sense a static structure. Computer software can function as dynamically morphing structures that directly affect system workings and computerized systems provide therefore a previously unmatched span of possible system behaviors for a single hardware design. In addition, the extreme quickness of electrodynamics makes it possible to alter software states and corresponding system properties at a speed that appears discrete to human beings. This may be contrasted against more traditional and mechanical kinds of systems such as the piston engine. Machine and functional properties for such a mechanical system are more intuitively understandable (albeit incomprehensibly complex in detail perhaps) because properties are rigidly confined to everyday physical relations such as motion and inertia, thereby enforcing a more continuous behavior. On top of this, there is usually a lot of observable output from mechanical machines, such as noise, exhausts, and vibrations that unveil some parts of the inner workings to their users, regardless of whether that is a designed goal.

Next characteristic of computerized technology is that software can contain complex and dynamic models or representations (C2) of virtually anything, and these dynamic models can affect almost all machine properties of these systems. This allows for, to begin with, in principle any level of complexity for automatic sequences of system actions (i.e., machine properties), which combined with the powers of computer logics make it possible to create almost any kind of system workings. Principally, software can provide regulation of machine properties at any level of abstraction, limited only by design creativity and programming skills. However, these models, or system-internal representations, are completely unobservable to human beings, unless there is explicitly designed feedback. Consequently, the machine properties of a system become incomprehensible for a human operator unaware of the current state of system-internal representations,

which is to be confused about what mode the system is working in. Mode-confusion is unfortunately a rather common problem (e.g., Sarter and Woods 1995). Furthermore, working with technological systems that comprise implemented models not completely compatible with reality, or with your understanding of things in the role as the responsible operator, is undoubtedly quite annoying, and sometimes even rather frustrating. It is so because working with automated systems resembles cooperating with human beings that have stubborn preconceptions or severe prejudices, but in this case these issues are even worse since these internalized ideas are significantly more persistent than the human kind and, in addition, impossible to assess because of the unobservable nature.[32]

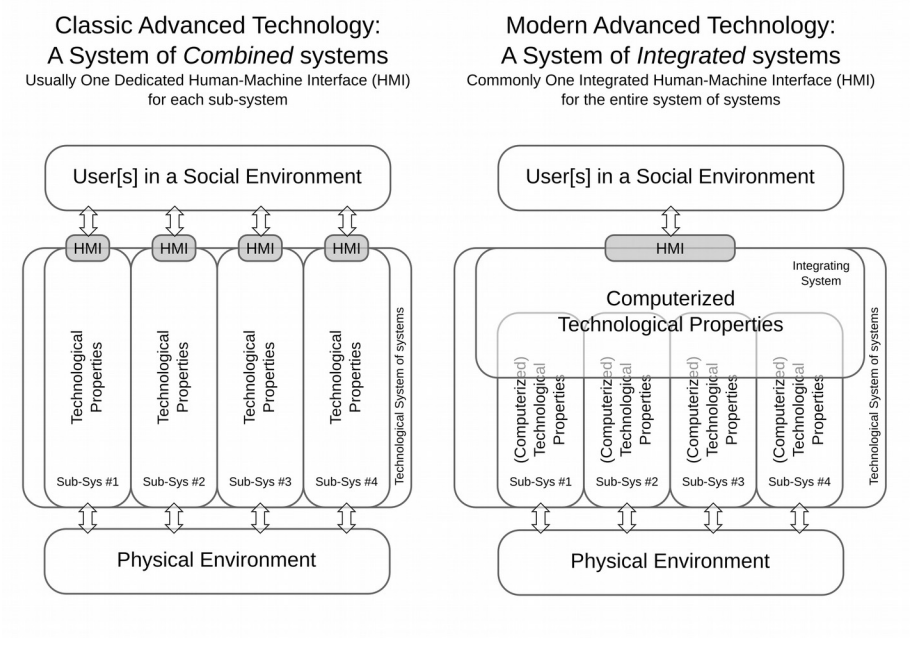


Figure 5.3: System integration before and present

The third characteristic is also a consequence of the software-oriented nature of computers. Electronically stored data is quite easily exchanged between computers, implying unprecedented possibilities and easiness for seamless systems-integration (C3) and for a virtually unbounded coupling of technological systems into systems of systems. While mechanical sub-systems in practice need to be physically bolted together and exchange of functionality in principle have to be made through axes and rods with rather obvious and often observable implications, computers can be combined more freely, even invisibly wireless, and communicate at virtually any level

[32]Try, for example, to convince an automation that it has gotten a few things wrong, and assess your persuasiveness!

of abstraction. For a human being this has the effect of erasing boundaries between sub-systems and of blurring system topology (Figure 5.3, right side), which may confuse human operators and reduce their system understanding (i.e., their system awareness, elaborated on in ch. 9.2), unless there is explicitly designed feedback to counteract the issue. The three characteristic properties (R1.1.3) of computerized technological systems are:

(C1) Completely unobservable and discrete dynamics

(C2) Complex and dynamic models affecting system properties

(C3) Seamless and arbitrary systems-integration

5.5 Summary of worldview models

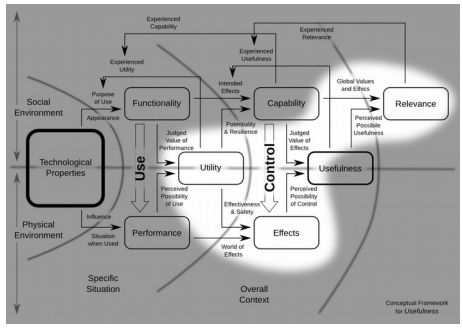
The presented worldview (R1.1.1, Figure 5.1) is considered to provide valuable means for analyses and for assessments of situated technological usefulness. It outlines for whom the technology of concern is to be useful as well as when and where it should be useful, ontologically. The separation of system environment into a physical and a social environment based on their fundamental differences in character is considered essential for the ability to distinguish different kinds of factors, aspects affecting system properties differently, for example, to distinguish factors determining possible effects from factors affecting system usage and intended effects, and from actual effects. The worldview introduces also the notion of a domain of effects distinct from the work domain, sometimes connected only by the system of concern, which is a relatively new kind of system design situation bringing specific demands on system designs that appear not yet fully comprehended.

The system properties model (R1.1.2, Figure 5.2 and P1-P3) supports analyses and assessments of usefulness by providing explicit means for describing what is to be useful. While this distinction between different kinds of technological properties perhaps is nothing new, the contributed layout is believed to highlight essential aspects often forgotten. In order to describe appropriately the character of the technological system of concern, it appears insufficient to speak only about concrete properties and well-known facts. Complexity can turn a concrete set of facts into an ambiguous mesh of interdependent workings and subjectively defined functional purposes thereby bringing abstract (social) aspects to the equation, which are fundamentally different properties compared to concrete physical aspects. Furthermore, all these properties depend on each other recursively.

The typology of computerized properties (R1.1.3, Figure 5.3 and C1-C3) provides, to begin with, additional means for describing what is to be useful (complementary to P1-P3), means required for appropriate descriptions of computerized technological systems of systems. In particular, it highlights

aspects that make computerized systems much more subjectively defined compared to mechanical systems because of the unique means for abstraction that computers facilitate. This integrating character affects functional properties, and to an even greater extent machine properties. The recursiveness in computerized systems can be an arbitrary mix of concrete hardware and abstract software implementations. It is, in particular, possible to develop system-internal abstraction layers in a manner bounded only by design creativity, which is why these systems get machine properties that are more subjectively defined than what is the case for traditional mechanical machines. While mechanical designs also have implicit design assumptions built into them, the abstraction structure for computerized system is less dependent on concrete matters. Abstracted and computerized system workings can take a more active part in the activities in which the system later is to be used (discussed further in ch. 6.3). These characteristic properties are particularly important in order to depict computerized abstraction layers and describe computerized automations.

6 Desired effects



The frameworks presented in chapter 5 consider the question about what technological system we are concerned with. They addressed the problem of describing appropriately the nature of technological properties and environments, when and where the designed technology should be useful, and for whom it should be useful. This chapter considers the question for what should the technology be useful, the purpose of the technology.

Precisely as for safety (introduced in ch. 1, elaborated on in ch. 7) and analogously for usefulness (introduced in ch. 1.4.5, explored in ch. 5, redefined in ch. 10), the idea that it is possible to state specifically what effects that are going to be desired in a future situation is here argued to be an unfortunate delusion. The idea seems based on an exaggerated faith in the existence of models and scenarios that are correct and complete descriptions of the world, thereby facilitating precise descriptions of desirable future effects. This idea is considered an exaggerated belief because all scenarios are, like all models, inherently wrong (e.g., Box 1976, Sterman 2002). Scenarios are always incomplete simplifications of a richer reality, which is perfectly in order because otherwise the models would be useless as their simplifications are what produce clarity and provide explanatory power. While a real future situation might follow the same principles and be similar to the scenario, it will never turn out exactly as the model. Reality will never be as clear-cut and comprehensible as a scenario, but instead richer and filled with subtle yet crucial nuances. Hence, the problem of predetermination of purposes lies in the specificity of their implementations, not in scenarios and models as such.

Desires, values, and meanings, on which purposes and intentions are grounded, are arguably to a great extent emergent within the actual situation thus not possible to determine unambiguously beforehand, not even if all conditions actually could be known in advance. Relevance, judgments about values and meanings, are contextual, "determined in the moment and in the doing" (Dourish 2004, p. 23). This inescapable uncertainty about future situations implies that the question about desired effects becomes not so much about what exact effects that are going to be desired, but more about why certain kinds of effects would be desired and how they will come to be. This thesis focus on how to generate situated useful effects because effects seems more likely to become desired if they can be sufficiently well tailored to the rich aspects of an actual situation, compared to potentially desirable predetermined effects that risk turn out irrelevant because they are applied according to stereotypical models. Predetermination requires a detached perspective seldom aligning completely with the involved perspective (cf. the plateau-society metaphor in ch. 1.4.3). The question about what will be desired must thereby be shifted slightly towards applicability of desirable technological properties (i.e., *can* they be used *for a specific purpose*?) and controllability (i.e., can they *be used* for a specific purpose?).

Some kind of modeled scenario seems required though, for an ability to select consciously what kind of properties that should be designed such that they can be applied under likely conditions and controlled to suit likely situations. There is no such thing as a universally useful technology because certain properties are by their very nature fundamentally incompatible with each other (e.g., spacious cargo hold and small footprint) and some properties can be truly useful for one purpose but highly counterproductive for another (e.g., strong but heavy). These contradictions imply that there has to be an idea about the purpose of each property, and this dyad, the purpose-property relation can be considered the designed utility of a system since the concept utility was found to mean "fitness for some purpose" (ch. 5). Of course, a system may have additional utility besides what is purposely designed that might be discovered by someone with a different interpretation of a situation and with other purposes in mind. The system may have certain contextual utility that largely might be unknown to the designer. Nevertheless, utility is yet another concept that cannot be defined unambiguously in advance and what effects the utility of a system actually will create depend on how such utilizable properties in fact become applied, which in turn depend on perception of how the properties can be used.

This chapter contributes a classification of two fundamental purpose-property dyads linked to each other in a tension called the character of utility, as well as a model for how effects come to be in principle. The classification facilitates the more in-depth discussion about applicability and controllability that follows in subsequent chapters. The theoretical framing is systems thinking approaches that in some sense reject the objectivity ideal

of, mainly, positivist philosophy of science and classic hard systems thinking. These approaches are summarized under the heading 'critical systems thinking and discourse' because without an attainable objective truth different viewpoints will forever imply a need for critical discourse that, presumably, will benefit from systems thinking.

6.1 Critical systems thinking and discourse

The inadequacy of classic (hard) systems thinking for psychological and social matters, and for value dependent aspects such as usefulness, may be considered connected to an exaggerated strive for objectivity favoring the detached perspective and predetermination. For operations research (OR), analyzing military and business operations thereby reaching beyond concrete physical matters into social aspects (sometimes also ethically difficult aspects), this strive for objectivity became a serious problem. Ackoff attributed the "pursuit of objectivity" as "the death of OR" (1979a, p. 93, both quotes), requiring "an alternative paradigm" (1979b, p. 189). Checkland called it a "divorce of theory from practice", arguing for a "debate about the adequacy of the paradigm of (natural) scientific objectivity" (1983, pp. 662, 663, respectively). Checkland further concluded that OR had "discovered *the logic of situations* ... [while ignoring that] any particular situation may be dominated by Blackett's 'chance events' and 'individual personalities'" (1983, p. 663, emphasis in original, brackets added) A situation that made OR suffer from the same fundamental weaknesses as systems analysis and systems engineering in which the logic of situations is used as a replacement for the situation itself.

Systems thinking approaches may be categorized as functionalist, interpretive, emancipatory, or postmodern, and critical systems thinking combined with the systems of systems approach aims to join the strengths of them all by taking a holistic perspective (Jackson 2000). For the present purpose, it suffices to explicate the technical-rational perspective of the engineering sciences from the perspective of the social sciences that regards actors as intentional agents (de Bruijn and Herder 2009) and identifies the clashing perspectives between the engineering and the human factors view (Kirwan 2000). This clash made Flood (1990) discuss liberating systems theory as a way to realize critical systems thinking (Jackson 1991, 2001, 2010) and it made Ulrich striving towards critical heuristics, critical systemic discourse, and critical pragmatism (1987, 2003, 2007).

In this context, what is important is the conclusion that regardless the theoretical possibility for a comprehensive understanding of the world thereby resulting in rationally objective values, in practice the inescapable presence of different viewpoints implies that there will be discourse, and there may be tensions between incompatible or even incomparable aspects.

Regardless the quality of arguments, if the discourse is transformed into a debate in which one side claims victory, essential insights may get lost. The notion of a tension is therefore used as an indicator of a rich problem. To begin with, the tension between purpose-property dyads in terms of different kinds of utility will be used to describe the character of utility.

6.2 Character of utility

The decision whether to use a certain technological system is often about judging benefits against problems because quite often, desirable system properties come with undesirable consequences. When designing a system, the problem is usually to strike a beneficial balance between counteracting aspects of system properties. Technological designs, especially non-trivial technologies such as integrated systems of systems, are always compromises, and system designs tend to become the pursuing of the perfect one. In short, the final properties of technological designs seem either be the result of a deliberate strive to achieve desired effects or the result of some focused efforts to avoid undesired effects. This idea of having the purpose of a technological system, or sub-system, fall within these two major categories ought to be uncontroversial. However, it appears that implications of the complex interdependencies between opposing purposes of system properties and effects often are underestimated. System functionalities tend to be considered well defined, giving the impression that there is a direct mapping between properties and effects. Consequently, some properties seem to guarantee safety (i.e., they are safe) and other properties seem obviously useful. In reality, however, the mapping between properties and subjectively defined values of effects is more complicated than that.

The purpose of the present common-sense kind of conceptual framework (R1.1.4) is to highlight the fact that system properties can be both supportive and competitive for a certain purpose at the same time. Systems design becomes thereby a question of balancing design efforts between competing purposes in order to create a desired kind of utility, rather than a question of selecting what specific design purposes to meet. The framework indicates that design is not about positioning an artifact on a scale, reaching from achieving desired effects to avoiding undesired ones. Rather, design is about choosing how to distribute design efforts on aspects from that scale. This character of utility distinction could perhaps benefit from being more elaborate, for instance by using a plethora of utility categories thereby avoiding the impression of two extremes of a single dimension, but that would also risk making it unmanageable. The basic idea here is that it suffices to have two categories to indicate a tension between different aspects (symbolized by the ↔ sign), in this case a tension within the character of utility. This tension is later used as one dimension in the two-by-

two matrix describing the character of usefulness as the contributed alternative definition of situated usefulness presented in chapter 10, a model addressing the main research objective.

To illustrate the tension between sought-after effects (i.e., tensions between the different categories of utility), the concept of safety seems once again relevant because utility and usefulness was found (in ch. 5) depending on each other, with safety as an obvious part of the equation. Safety is an essential aspect of usefulness, simply because discovering that a technology is unsafe and dangerous will quickly make it be considered unworthy the risks, thereby rendering it practically useless, unless, of course, the advantages from using it clearly outweighs the risks. That is, lack of safety can reduce the usefulness, but the reverse relation does not hold. The presence of safety does not ensure usefulness. The missing concept in this equation is effectiveness (discussed further in ch. 7.3). For a system to be useful in practice, it cannot only have properties providing means for a reasonably safe operation but must also have effective properties. At the same time, a useful system cannot have effective properties only but must also provide means for a reasonably safe operation.

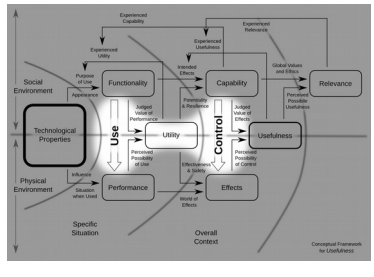
Safety and effectiveness stand thereby somewhat in opposition to each other, linked by the concept of usefulness. An otherwise effective system can have its usefulness reduced by a lack of safety, and a safe system can have its usefulness reduced by a lack of effectiveness. To be truly useful, the system must have properties that provide both effectiveness and safety. Hence, there is a tension between having utility in the sense of providing effectiveness by achieving desired effects and having utility in the sense of providing safety by avoiding undesired effects, a tension forming the character of utility of the technological system of concern.

The character of utility
Achieve desired effects ↔ Avoid undesired effects

So, the ideal design strategy would then simply be to maximize both safety and effectiveness, thereby creating a system of perfect usefulness, would it not? It seems, however, there is another incompatibility between these two aspects. While both kinds of utility are desired (i.e., sough-after purpose-property dyads), they may depend on the same set of physical properties, thereby making the subjective purposes clash. Effective properties are quite often dangerous properties, and measures to increase safety are therefore likely to hamper the effectiveness. The purpose is what matters. For example, the purpose of a chainsaw is to cut trees and thick branches and, inescapably, the very properties that make the machine effective for that purpose are in addition effective for damaging human limbs. Design solutions intended to make it difficult to cut accidentally human limbs, risk thereby also reduce branch cutting effectiveness. Purposely designed

effective properties become, however, mostly not unsafe until they end up being used unskillfully, without proper system knowledge, or for other purposes than those governing the design.[33] Safety measures tend therefore to confine usage to precisely those situations and purposes for which the system was explicitly designed. This connects to the rhetoric theme of this thesis, the undesirable consequences of predetermination, and the difference between a calculative approach and a generative approach to usefulness. While safety measures that confine usage to predetermined purposes well enough may eliminate certain undesired effects related to malign purposes identified in advance, they tend also to eliminate the achievement of desired effects related to benign purposes not identified in advance. Essentially, the result is a confinement to predefined purposes, determined by detached models and calculative values. It appears thereby necessary to distinguish between calculative effectiveness and generative effectiveness, as well as between calculative and generative safety. This issue will be addressed in more detail in chapters 7.3 and 7.4.

6.3 From conditions to effects, in theory



How do effects from technological systems come to be? The answer might be thoroughly complicated though one thing is certain. Technological artifacts do not apply their properties by themselves. There are always one or more human beings involved when effects from technology occur. This is the case because even the most self-governing of systems, a so-called autonomous and artificially intelligent system capable of acting on its own principally without any human involvement, is still initially designed, built, and has at some time been switched on by human beings, hopefully with sufficient insights about what this might lead to. Even a stupid technological artifact that create effects simply by existing in a situation depend on an

[33]Rather often, these other purposes (especially certain genuinely evil purposes) are very far-fetched for a system designer primarily concerned with trying to do well and achieve desired effects. It is, for example, quite understandable that a green-laser-pointing-device developer fails to realize the potential to use it to blind airliner pilots during flight (e.g., on the final approach to landing) thereby risk causing disasters, a scenario that unfortunately has become almost commonplace.

initial human activity because its effects will not come to be unless someone has applied the artifact as part of the conditions for the situation in which its properties create effects. That is, for technology, there has to be some kind of human activity for effects to emerge from conditions (R1.1.5).

Human activities are in turn governed by psychological and physiological aspects because without an incentive to do something and without an ability to do so there will be no activity. Exceptions to the requirement of incentives would be our reflexes and subconscious behaviors that make us do things without really having a conscious incentive, but for reflexes and behaviors to be relevant when considering control of technological systems, there must arguably first have been a conscious decision to get hold of the technology and put it into operation. That is, when considering effects from technological systems it is the psychological domain and the activity domain that constitute the bridge from conditions to effects (Figure 6.1).

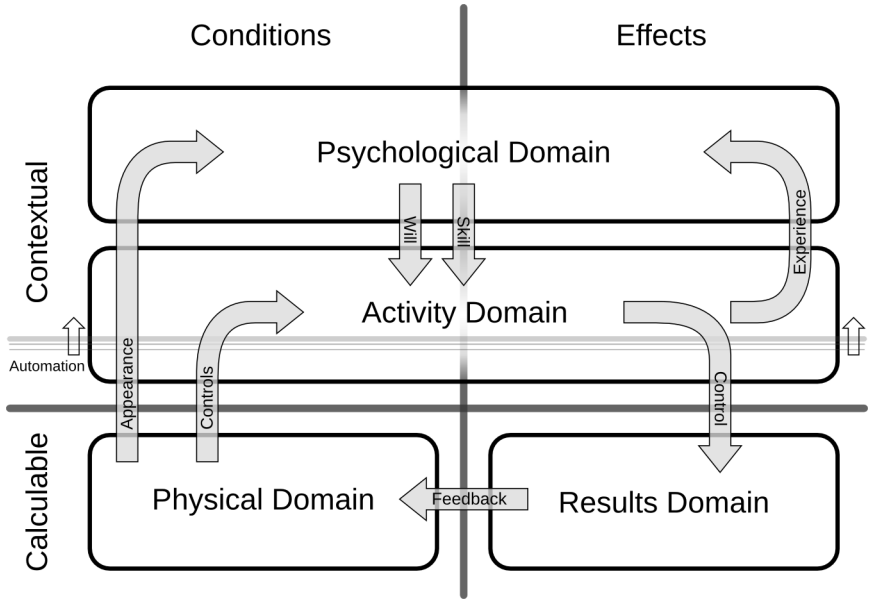


Figure 6.1: From conditions to effects (R1.1.5), in theory

The psychological domain is where incentives for use emerge, regardless whether they are conscious or subconscious incentives, and the psychological domain is in addition where the evaluation of results takes place, an evaluation that also may be both conscious and subconscious. These two psychological processes rather different in character (i.e., conscious and subconscious thinking) seems characteristic for human beings and are addressed by the notions of controlled and automatic information processing (Shiffrin and Schneider 1977a, 1977b, Birnboim 2003a), by

system 1 and system 2 (Stanovich and West 2000, Kahneman 2003), or thinking fast and slow (Kahneman 2011). The two modes are in some sense also addressed by the knowing-that and knowing-how distinction (Ryle 1945, Dreyfus and Dreyfus 1980, Dreyfus 1992) and by the distinction between explicit and tacit knowledge (Polanyi 1966). For the present purpose, it suffices to be aware of the different characteristics of human psychology on a principal level.

The activity domain is what actually is taking place. The process can be described as a never-ending loop of continuously evolving situations, driven by human activities and closed by physical feedback of concrete results that alter the conditions for current and future activities, and by psychological feedback in the form of experiences that affect interpretations and incentives as well as subconscious intuitions and feelings. The psychological and activity domains appear, however, extensively intertwined because we human beings are continuously and simultaneously both physical and intellectual creatures, regardless of how much effort we put in trying to separate these two aspects. This interdependence shows for instance as that incentives (i.e., will) tend to develop from experiences of similar situations and that they tend to be quite influenced by confidence in the ability (i.e., skill) to go through with certain activities. Skills are also thoroughly dependent on having gone through with such activities before, which naturally require once having had the will (i.e., the incentive) to do so. Obviously, this means that at some early stage in training, the incentive must be strong enough to overrule any doubts of ability for going through with the activity such that experience can be gained and confidence in the ability achieved.[34]

This loop-of-life, the constantly ongoing exchange between the mental world (i.e., the psychological domain) and the physical world in which the human activity domain is thoroughly rooted, is probably one explanation for why our thoughts and behaviors cannot be fully understood without considering the overall context. We are always situated in a larger context than whatever specific situation that is considered, affecting the interpretation of the local context. This is also one of the fundamental differences between human beings and technological systems. The technological system will work the same in two situations distinct in time, in two different contexts, if those specific conditions relevant for its operation are the same. Human beings will not work the same, even if conditions are the same! Besides being inescapably variable in physical behavior, during the timeframe that makes the situations distinct, the human being will have gained new experiences adjusting interpretations and incentives. The plethora of social aspects that makes up the overall social context evolve

[34]This fact is perhaps a plausible defense of the tendency among people to overrate their own abilities (e.g., Kruger and Dunning 1999), an attitude that actually seems required for ever beginning to develop new skills.

over time, at different rates and with complex interdependencies. Because of this interdependency, the resulting effects for each of all these social aspects are included in the conditions for next iteration making the activity-psychology bridge between conditions and effects eternal and continuously smooth. It has in practice no beginning and no end (except, perhaps, at birth and death for each involved individual), and the systemic notion of iterations becomes somewhat irrelevant for contextual aspects. Rather, the context should be considered a continuously evolving and constantly morphing object impossible to define exhaustively without taking all the involved human beings entire lives into account, and perhaps some inherited aspects. This unbounded continuity may be regarded as a significant characteristic of social systems and a reason why the notion system might be inappropriate for social matters (the question was also raised in ch. 5.1).[35]

The properties of the technological system of concern constitute, naturally, much of the conditions for an activity that intend to use that particular system to achieve some desirable effects, and the characteristics of these properties have therefore great impact on the activity. This impact is particularly evident for machine properties (P3, ch. 5.3, and C1-C3, ch. 5.4). Arguably, the significantly larger influence from machine properties on activities, compared to the influence coming from functional (P2) and physical (P1) properties is due to the fact that dynamic system workings take a much more active part in activities, compared to physical and functional properties. Despite the possibility to interpret their functionality locally, physical and functional properties are still passive and static, simply providing conditions for use by virtue of what they are. The artifacts have their properties and it is entirely up to human beings to make use of them in their purposeful activities. Dynamic machine workings become on the other hand a significant part of the activity itself, and sufficiently advanced automations can principally govern activities all by themselves, once they have been triggered.

All this has the consequence that the more complex machine properties, the more they interfere directly with activities otherwise governed by human beings, and machine properties are principally determined in advance. This predetermination of activities, by automation and complex computerized machine workings, may therefore be described as raising the boundary between the calculable and the contextual into the activity domain (Figure 6.1). Automation does, so to speak, annex a piece of the context and make it calculative by active intervention. The three characteristic properties of

[35]If one should try a mathematical metaphor for the incommensurable nature between the calculable physical world and the contextual social world, the Fourier transform might be appropriate. The concrete physical world could then be described as a function in the time domain, while the abstract contextual world would be in the frequency domain, and in order to get from time to frequency one must integrate all aspects from $-\infty$ to $+\infty$. The social aspects (the frequencies) are emergent within the system as a whole in the sense of being defined by every event from the beginning of time until eternity.

computers (C1-C3, ch. 5.4) add further to this problem because the incomprehensibility that follows from discrete dynamics and covert workings make it even more difficult for people to understand system workings and develop incentives for trying to control it to work differently. The result is that human activities become calculative as well, and context loses some of its situated richness.

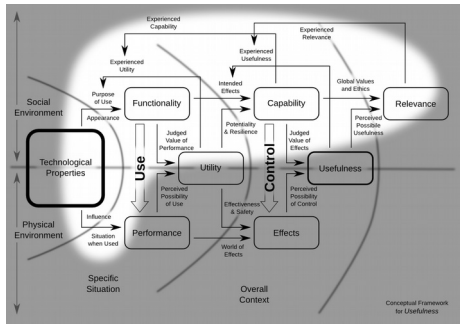
6.4 Summary of value models

The tension within the character of utility (R1.1.4) indicates the ubiquitous compromising associated with technological designs, a compromising that comes from incompatible physical properties and conflicting values. The latter is, arguably, the more complex issue, especially for advanced systems of systems. One essential aspect of design compromises is that they depend on socially situated values and purposes. System properties, regardless whether they are physical, functional, or machine properties, are selected for a purpose, which is to achieve certain effects. However, what effects that will be achieved depend on how the properties in fact are applied, which in turn depend on values and purposes within the social context of use.

The model (the map) of how conditions become effects in theory (R1.1.5) is therefore a necessary foundation for further scrutiny of situated usefulness. The model illustrates the connection between activity and context, between will, skill, and actual control, a connection also crucial for understanding the impact of automation on value judgments and purposes.

With these two models in conjunction it is possible to see the impact on human autonomy that follows from extensive automation, a design approach enforcing a detached kind of system control (discussed further in chapters 10.3 and 15). The more complex the system of systems becomes, and (computerized) automation can facilitate a high level of complexity, the more purposes and values become implicitly assumed and implemented. Given the premise that people (e.g., system operators) understand the world and develop incentives much by situated interaction with it, they are likely to have their desires align with predetermined purposes because automated system control creates an interaction governed by these purposes. Such a development has the effect of locking purposes of use to those purposes that were identified during design, which thereby assumes that values governing these purposes are objectively correct and valid in the case of use as well. Essentially, this is to enforce a heteronomous behavior according to calculative detached values.

7 Uncertainty and unpredictability



In chapter 5, frameworks and models were presented for describing what we are discussing, what properties we are considering, where and when and by whom these systems are to be used. Chapter 6 contributed frameworks suitable for describing why the systems are used and for describing the character of utility based on the purpose for which system properties have been designed. There was also a basic map drawn, describing the principle (i.e., the theoretical) path from conditions to effects. These frameworks are, however, all in some sense ontological and detached as they aim to assist in describing what things are. This chapter, and the subsequent two, focuses on how to get where we want. We are now about to start walking the map from detached conditions to situated effects.

To begin with, this chapter will focus on obstacles for, and difficulties in, actually achieving desired effects. This is because not all obstacles can be taken care of in advance, the future is always to some extent unpredictable, and different uncertainty characteristics seem to require different strategies. The present stance is that it is virtually or practically impossible to have a complete understanding of everything relevant in advance, especially everything relevant for non-trivial matters involving human beings with competing interests. Focus falls then naturally on why there is uncertainty, why things are unpredictable, and on how these uncertainties manifest themselves. To help sort out the reasons for unpredictability a framework for describing the character of uncertainty is contributed. It is a framework supposedly useful for assessing and describing what means human operators may require maintaining sufficient control. The uncertainty framework is followed by an elaboration of the discussion initiated above, about

calculative and generative effectiveness and safety, an elaboration presenting two additional tensions, similar to that within the character of utility. Because a natural consequence of uncertainty and unpredictable situations is unsafe operation of systems, the theoretical framing is safety science.

7.1 Safety science

One undesired consequence of uncertainty and unpredictability is the problem of preparing for and avoiding hazards. Real uncertainties oppose explicit safety preparations because it is impossible to safeguard against the unpredictable with safety-measures that depend on explicit predictions of conditions and events, which for example is the case for many technological safety systems. With a hard systems thinking view (i.e., a detached perspective assuming deterministic system workings), uncertainty becomes the cause of hazards. To achieve safety under such assumptions, to avoid undesired effects with an essentially deterministic worldview, becomes therefore to eliminate uncertainties in order to facilitate comprehensive preparations such that all hazards are neutralized. Artificial elimination of uncertainties (e.g., by technological means) often takes the form of explicit enforcement of predictable system behaviors. The problem with this approach is that technology can only enforce the behavior of certain limited sub-systems, predominantly physical machinery, but not on more complex real-world events and socially defined aspects. This problem is presumably a major reason for the safety paradoxes described by James Reason (2000). Explicit implementations of artificial predictability cause systems to break down because necessary adjustments become impossible, thereby making disasters man-made (Turner and Pidgeon 1997, Gherardi *et al.* 1999).

While technology does influence human behavior, as described, for example, by the theory of technological structuralization (Orlikowski 1992, drawing upon, Giddens 1984), technological systems do not enforce a completely predictable behavior on the social system with which the physical systems interact. Deliberate choices and intentions grounded in social norms and values do, however, often imply control input of higher orders (Parker *et al.* 1992, Brehmer 1994) resulting in limited control possibilities on lower (technical) levels thereby exacerbating consequences of such higher order decisions. An abstracted and simplified control model limits control possibilities, the essence of the law of requisite variety (Ashby 1956). The calculative mindset tend on top of this to come to the conclusion that consequences of the limited control possibilities are human errors (Reason 1990), thus addressing the problem by imposing further limitations to mitigate consequences of slips, lapses, and mistakes. However, the natural curiosity of human beings and, from a calculative perspective, the irrational diverging from optimal routes is in fact a necessary condition for discovery

of novel hazards and application of heroic recoveries (Reason 2008). Humans have, so to speak, a natural tendency to continuously stress test controlled systems, which arguably is required for generating edge awareness, which require that system designs allow such testing without breaking, and that the designs provide feedback about performance limits. These two sides of safety lies at the core of the four safety paradoxes identified by Reason (2000, p. 4), which are (quote):

- Safety is defined and measured more by its absence than by its presence.
- Measures designed to enhance a system's safety – defences, barriers and safeguards – can also bring about its destruction.
- Many, if not most, engineering-based organisations believe that safety is best achieved through a predetermined consistency of their processes and behaviours, but it is the uniquely human ability to vary and adapt actions to suit local conditions that preserves safety in a dynamic and uncertain world.
- An unquestioning belief in the attainability of absolute safety (zero accidents or target zero) can seriously impede the achievement of realisable safety goals.

The book "Normal Accidents – Living with High-Risk Technologies" by Charles Perrow (1999)[36] has survived and continued to thrive because "it links, it frames, it provokes" (Weick 2004, p. 27). The most provocative aspect is perhaps that the normal accident theory, which evolved from this book, indicates that the present paradigm of technical rationality cannot ensure safety and, rather, is causing disasters. Because nobody likes to know that they have understood something wrong, "this must be at least occasionally irritating to engineers and designers" (Kirwan 2000, p. 664). Based on centuries of conflicts, it is possible to go further and state that the engineering approach (to command and control) "create inefficiency (instead of efficiency), unpredictability (instead of predictability), incalculability (instead of calculability) and complete loss of control … the antithetical problems, ironies and productivity paradoxes that, when all else fails, fall into the lap of ergonomics" (Walker *et al.* 2008, p. 483).

Such normal disasters are systems accidents, a term actually preferred before normal accidents by Perrow (2004), disasters that are bound to happen because of the constitution of the system. The normal accident theory considers the social side of technological risk and centers on two

[36]The original book was published in 1984. The reference used in this work is a reprint with a quite extensive addendum (58 pages) briefly scrutinizing several noteworthy accidents that occurred just after the original publishing of the book. Consequently, the timing of the book was academically fortuitous, a fact that Perrow (2004, p. 10) holds as a major reason for its great impact and the fact that "people declared that [he] had bragging rights". The accidents that followed were the Union Carbide chemical plant disaster in Bhopal in 1984, the Chernobyl disaster in 1986, and the Challenger space shuttle accident also in 1986.

identified dimensions of risk – complex versus linear interactions and tight versus loose coupling. Systems bound to fail are those with many non-linear interactions (i.e., systems of high interactive complexity) that also are tightly coupled. One problem with the conventional engineering approach to safety is that it tend to imply that "we load our complex systems with safety devices in the form of buffers, redundancies, circuit breakers, alarms, bells, and whistles" (Perrow 1999, p. 356) only making the systems more complex and tightly coupled as "these systems are not independent of one another. The alarm rattles the bell; the bell shatters the whistle; the whistle explodes; and suddenly the whole system collapses" (Sagan 2004a, p. 17). The problem is neatly summarized as the problem of redundancy problem (Sagan 2004b). At Chernobyl, tests of a new safety system helped produce the meltdown and the subsequent fire (Perrow 1999). Furthermore, more and more of our systems are becoming implemented in software that inescapably apply the engineering view on matters, and the gravity of this problem can be illustrated by the fact that already in 1981 more than 80 percent of our weapon systems required computer software (Leveson 1986, p. 125), a figure that today presumably is much higher.

System designs and software implementations are always implemented beforehand, which means that the character of uncertainty is at the core of the predictability problem, "the completeness and accuracy of the model for the type of system being considered will be critical in how effective the engineering approaches based on it are" (Leveson 2004a, p. 66). Accidents occur when uncertainties are addressed with improper means and the means to address uncertainties are, presumably, dependent on the reason for their existence. Therefore, a list of presumed plausible causes of uncertainty will serve as a conceptual framework useful for describing the character of control required to achieve situated usefulness.

Kirschenbaum (2002) lists three major causes of uncertainty for the context of automated systems: observation uncertainty (e.g., signal/sensor error, transmission loss or distortion, data age, etc.), processing uncertainty (e.g., statistical estimation, aggregation, misapplication of assumptions, rule conflicts [in AI methods], etc.), and reporting uncertainty (e.g., misrepresentation of uncertainty, communication failures, poor display, etc.). Olson and Olson (2000) add that distance matters as a direct consequence of interdependent and intertwined social aspects. There are, essentially, two approaches to address the problem of uncertainty, either by accepting the normal accident theory (NAT) or by opposing it and treat safety as equal to reliability (i.e., predictability) and strive for high reliability organizations (HRO) (Leveson 2004a, 2004b). Or, perhaps is there a third way, the systems-theoretic approach to safety (Leveson *et al.* 2009), clarifying the strengths and weaknesses of NAT and HRO respectively as well as taking the technological change and the corresponding changing nature of accidents into account. The changing nature introduces new types of accidents, a

decreasing tolerance for single accidents, more complex relations between humans and automation, and a changing regulatory and public view on safety. In the end, this is all about safety cultures (Reason 2000, Westrum 2004b, Parker *et al.* 2006) and safety research has recently begun to extend the concept of safety to that of resilience (Hollnagel *et al.* 2006). However, applications of probabilistic risk assessment (PRA), high reliability analysis (HRA), and the resilience approach, have shortcomings (Sheridan 2008).

7.2 Character of uncertainty

The following notions (R1.2.1) describing the character of uncertainty should be considered very general, they are supposed to be applicable virtually everywhere in a sociotechnical system, and at any level of abstraction. Actually, they are deliberately abstract in order to be as holistic as possible. Whenever used, it is therefore crucial to be specific about what system or activity that is referred to, not least because the recursive nature of systems thinking tend to mix these aspects up. For example, a complex technological system may have a large scope by itself, or maybe the activity in which the technology is merely an element is what makes up the scope. This means that the following notions of uncertainty require initially to be related to the outline of the overall sociotechnical system (R1.1.1, ch. 5.2) and to the respective characteristic properties of its components (R1.1.2 ch. 5.3 & R1.1.3 ch. 5.4). The character of uncertainty may then be described by use of the following characteristics:

Scope – size of state space: The total number of relevant factors and the range of each factor form the state space for the entire system. It is difficult to grasp a large state space, and thereby to explain or predict its behavior, even if the system is assumed to be governed by well-known laws of interaction (i.e., by formally describable dependencies between factors). Scope may of course be seen as one significant factor of complexity, the next character of uncertainty, but it is here believed that scope and complexity are not necessarily coupled. Specifically, a simple system (i.e., with low complexity) may still be quite unpredictable simply due to an overwhelming size of the state space.

Complexity – incomprehensibility: Even if the system essentially is confined and thus may have a comprehensible scope, it may be complex enough to be thoroughly unpredictable. Complicated dependencies, discrete, non-linear, circular, or multiple interacting dependencies create complexity. Contextual dependencies are specifically difficult to assess, which is an issue that may create incomprehensibility. Situations are stretched over a greater timeframe when having contextual dependencies, which also implies dynamically changing dependencies. In terms of hard systems thinking and material physics, complexity may be exemplified as the kind of contextual

dependencies that make past state trajectories relevant for the current state, as for hysteresis when bending shape-memory alloys.

Dynamic range – possible change in tempo: By dynamic range it is meant the possible difference in tempo, in how quickly things may be changing. The work situation for intercontinental airline pilots and the difference in work intensity during takeoff and landing compared to on-route flight is one example of a rather large dynamic range. Dynamic range may create problems since it can be difficult for both technology and humans to be well calibrated for the entire range. For example, a system design may focus on one end of the dynamic scale and thus induce floor or ceiling effects.

Dynamic agility – possible rate of change in tempo: If dynamic range is the largest possible difference in tempo then dynamic agility is how quickly the system may change between its extremes in tempo. The work situation for airliner pilots (e.g., the Air France case, ch. 12.3) is once again a good example. The low-intensity activities during on-route flight are changed very quickly into an extreme tempo during a severe incident.

Rigidity – extent to which constraints may be assumed being obeyed: This property, cause of uncertainty, or salient character of uncertainty is perhaps also one significant factor of complexity. It is, however, believed to be a thoroughly expressive notion by itself, perhaps particularly for social factors. To begin with, a concrete physical example, the concept of a roadblock can be used. Roadblocks clearly indicates a constraint, one should not go that way. Constraints, however, may be differently rigid. It is less likely that someone will ignore the block and simply run through a large concrete stopper, compared to a thin wooden gate. That is, it is more certain that the road behind the concrete stopper is free from undesired traffic than behind the wooden gate. Hence, a system with rigid factors is less unpredictable than a system with flexible and dynamic factors. Social factors are seldom unquestionably rigid and, in particular, argued to vary substantially in rigidity. For instance, constitutional laws are assumed more rigid than informal social codes of conduct, although they are both abstract and merely social arrangements. In particular, social factors may affect the rigidity of other social factors. The practical rigidity of a thin wooden gate is for example greatly determined by the social environment. That is, rigidity is to what extent factors may be assumed to be valid, how enforcing they are. Physical laws tend to be more rigid, hence more often valid, compared to social constructs. Alternative but opposite concepts to rigidity could be rejectabilty or ignorability.

Distance – separation of the work domain and the domain of effects: Spatial distance is obviously a significant factor that separates the work domain from the domain of effects. Concrete examples already mentioned are teleoperated UAV systems and Internet applications. There may also be a temporal distance, sometimes because of spatial distance, where effects from operating the distant technological system appear significantly later and

become observable at the controlling end at an even later moment. Temporal and spatial distances are also often common sources for concrete uncertainties like measurement errors, prediction errors, transfer errors, and so forth. Spatial and temporal distance may separately, or in combination, imply a completely different context as well, which means that it should be possible also to speak about a contextual distance. When human beings feel dissociated or detached, for instance when being physically, temporally, or contextually alienated, they tend to lose engagement as well, which implies that it should be possible to speak about an emotional distance as well. That is, the distance between the work domain and the domain of effects may be described by being made up of a combination of these four different kinds of distances. How much separated, and in what way the work and effects domains are separated, are considered significant characteristics for a sociotechnical system. That is, without a separation of domains the operators have access to all perceivable kinds of information, and particularly to direct natural feedback from the effects of using the technological system. With significantly separated domains, the operators are confined to information that designers thought of having the systems convey, or to information available from other systems. Direct feedback by proximity and the presence of various natural analog effects are supposedly important for human beings for the ability to 'get the hang' of things and develop an appropriate gut feeling. Conversely, a lack of direct feedback seems to create a sense of alienation and detachment, perhaps leading to less engagement. Hence, natural and direct feedback appears important for human system operators to avoid become contextually and emotionally detached. The uncertainty framework is summarized as the following bullets:

(U1) *Scope*: The larger the system of concern, i.e., the less a problem of simplicity, the more uncertainty because it implies more openness from missed or disregarded factors and misrepresented relations.

(U2) *Complexity*: Complicated, multiple, circular, discrete, and non-linear, etc., dependencies cause uncertainty despite well-known laws of interaction because of non-deterministic properties.

(U3) *Dynamic range*: The larger the difference in possible tempo, the more difficult it becomes to have the ability to calibrate controls for the entire range, without which the outcome is unpredictable.

(U4) *Dynamic agility*: The quicker possible change in tempo is, the more difficult it becomes to calibrate controlling efforts quickly enough, without which the outcome is unpredictable.

(U5) *Lack of rigidity*: The less rigid the rules for system workings the more difficult it becomes to predict, because it is less certain that the system actually will work according to the rules.

(U6) *Distance*: The looser coupling between the work domain and the domain of effects, the more unpredictable system workings and outcomes become, and the separation can be of different character based on its cause.

(U6.1) *Spatial distance*: The longer the physical distance, the more uncertain feedback there is because spatial distance increases the probability for transmission errors such as lag or distortion

(U6.2) *Temporal distance*: The larger the lag between control and effects, the more difficult it is to apply appropriate control measures, especially with demanding dynamics.

(U6.3) *Contextual distance*: The larger the spatial and temporal distance the more distinct becomes the control situation from the where effects occur, implying different interpretations and meanings resulting in unpredictable effects.

(U6.4) *Emotional distance*: The larger the contextual distance the more emotionally detached human operators tend to become and thereby less attentive to feedback, which makes control measures less appropriate and distant effects surprising and unpredictable.

7.3 Effectiveness and safety

Without uncertainties, situations would be perfectly predictable and there would be no need for situated control of system behavior. Certain derivable properties would with certainty lead to well-known effects. The concept of situated usefulness would in that case be obsolete (i.e., irrelevant) because system effects could then righteously be predetermined as desired. However, in particular for socially dependent aspects, there are always uncertainties, presumably uncertainties describable by the character of uncertainty framework above. Every non-trivial real-world situation is unpredictable to some extent, making situated system control required if effects are to align with intentions. Control must be made according to local and unpredicted variations of influential conditions such that system behavior is kept aligned with desirable purposes despite the local variations. However, purposes are possibly contradictory even for a simple system, a relation characterized by the tension within the character of utility, and because purposes largely change with changing conditions, the desired balance for this tension may have to be reconsidered for novel situations. These desires must be reconsidered within the novel situation itself because that is where the conditions have a meaning. The system must be controlled not only

according to local physical conditions but also to suit local purposes, for otherwise the effectiveness is irrelevant. Furthermore, the system must be controlled such that undesired effects are avoided despite changing conditions, meaning that safety must be maintained, to avoid having consequences overrule benefits. For a system to be truly useful, it must provide both effectiveness and safety, in unpredictable situations as well. For a system to be truly useful, as in having situated usefulness, some kind of situated controllability must exist because the world is dynamic and unpredictable. Sometimes the requirement for safety may require complete system shutdown (i.e., to abandon the striving to achieve desired effects) and, sometimes, the desire for effectiveness may require going through with system operation at the expense of safety (e.g., for high-risk enterprises or when collateral damage is deemed acceptable). Depending on system properties, on the desired character of utility, and on the character of uncertainty, the requirements for situated control differs.

Whereas the presence of situated conditions by definition requires situated control input, conditions for control are still largely determined in advance, by systems design. What are controlled in a real situation are the properties of the technology, and what can be controlled depends on system design. Control can be made either through explicitly implemented controls (i.e., system input channels present in the human-machine interface) or through manipulation of system environment (e.g., pull the plug to stop a non-responding system), or by combinations of both. System control can also be automated. Automation may provide what at first might be considered a kind of situated control, by being implemented to take certain situational parameters into account, parameters for which sensors exist or parameters possible to load in advance as on-board data. This kind of control is, however, not really situated because how the situational parameters are taken care of and what situational aspects that in fact are considered, are determined in advance by system designers and programming.

Borrowing notions from the different safety cultures discussed above, a calculative design culture relies on the predictability of situations while a generative design culture does not. The resulting designs will differ in character, depending on the prevailing design culture. Calculative designs assume that predetermined automations can be both effective and safe, while generative designs acknowledge the need for situated control in order to facilitate contextually relevant effectiveness and safety. It seems there is a contradiction between these cultures similar to the contradiction between effectiveness and safety, which is why the approach to controllability, the case of design is presented as another tension:

The approach to controllability, the case of design
Calculative design culture ↔ Generative design culture

To describe this tension between a calculative and a generative design culture, both effectiveness and safety is taken as notions of a calculative approach, to be contrasted against corresponding generative notions yet to be presented. The word effective means "producing a decided, decisive, or desired effect",[37] making the concept of effectiveness have a rather straightforward calculative meaning in terms of to what extent these decided, decisive, or desired effects are achieved. With well-defined goals and parameterized definitions of desired effects, the notion 'to what extent' allows for an after-the-fact calculation of ratios between optimal and achieved effects. On the other hand, an involved perspective might reveal other relevant parameters than the modeled ones, and perhaps in addition other relevant goals against which the effectiveness should be evaluated. Situated or involved aspects are in some sense to calculative or detached aspects what future situations are to present observations. They may include parameters not yet observable or not yet discovered as being relevant. The generative aspects appear therefore not possible to describe by the same means as calculative aspects, yet they seem connected. Hence, the calculative connotations of effectiveness make it necessary to find, or invent, a supplementary generative concept. For the lack of alternatives, potentiality is tentatively suggested for this purpose (ch. 7.4).

The considering of safety as a calculative notion might be slightly more controversial than designating effectiveness as calculative. In fact, the concept of safety is in part selected for rhetoric reasons. In is intended to function as a discursive provocative statement, about the contemporary shift in concept meanings (elaborated on in ch. 15.1). Safety is also selected as the calculative notion for avoiding undesired effects because within safety research there is already an established contrasting term, *resilience*, which seems to advocate the generative dimension. In order to explore further the tension between the calculative and generative culture, the concepts of potentiality and resilience must first be scrutinized.

7.4 Potentiality and resilience

The difference between safety and resilience may be described as that safety denotes a calculative (i.e., detached and static) being safe, while resilience describes a generative (i.e., involved and dynamic) becoming safe. To become safe can mean both to avoid actively becoming unsafe and to return to safety if somehow having become unsafe. Resilience engineering is characterized as striving to design systems in which it is possible to ensure distance to, oppose closing-in on, prepare for anyway ending-up facing, and provide measures to recover from, hazards in general (Hollnagel *et al.* 2006). Analogously, effectiveness or efficacy may be described as a calculative

[37] http://www.merriam-webster.com/dictionary/effective

(i.e., detached and static) being efficacious and potentiality as a generative (i.e., involved and dynamic) becoming efficacious. Designing for potentiality would then be to strive for systems in which it is possible to get close to, explicitly close in on, prepare for facing, and provide measures to utilize, opportunities in general. Potentiality and resilience are, arguably, both about the possibility to seize the moment, take opportunity of rare occasions, work the flow, tip the balance by a feather touch with accurate timing, and so on, something that obviously require situated control. However, focusing on calculative effectiveness and safety tend to inhibit generative application. Hence, the character of controllability, the design result, has a tension within.

The character of controllability, the design result
Predetermination ↔ Situated control

With well-defined goals and a calculative culture assuming well-defined problem-spaces, how to achieve these goals becomes a matter of finding and following an optimal path towards the goal. The inherent variability and unpredictability of real-world situations make it, however, practically impossible to derive such optimal routes to follow. Most important, though, is the fact that many goals cannot be precisely defined in advance as they depend on contextual aspects only available within a specific situation. Unpredicted drifts away from a predefined safe route are from a calculative safety perspective always hazardous because it means getting closer than prescribed to known dangers, or farther away from known safety, but there is obviously also a possibility for the opposite. What if there happens to be an unpredicted drift towards a unique opportunity, meaning that a new goal requiring a different path has emerged? Clearly, a truly useful design must allow seizing such opportunities, which is why effectiveness is insufficient and potentiality is required as well. Furthermore, unpredicted drifts away from the predefined route may be required for discovering either that the predetermined goal or path is inappropriate due to unique context specific aspects, which is why safety analogously is insufficient requiring resilience as well. Potentiality and resilience can roughly be described as the providing of unspecified utility, a technological assistance in the discovery of unanticipated alternatives of action, and a general support for seizure of such unpredicted moments.

In terms of car driving and edge awareness, a stability control technology cannot always maintain stability unless it also is given authority to control direction and speed, simply because of the laws of physics (e.g., Newton's second law applied to friction forces and vehicle momentum) and the law of requisite variety (Ashby 1956). Thus, in order to maintain stability the control technologies have to interfere with the ultimate responsibility of the driver, which is to maintain continuously a judicious control of direction and

speed. The interfering is often executed by imperceptibly impose limits on, or slightly take over, vehicle control according to implemented performance models (e.g., by reducing engine power, apply brakes, etc.). While having the benign effect of keeping the car within its driving envelope,[38] it has also the malign effect of making the driver unaware of being close to the real envelope edge that is modeled by the assisting system. The problem with this approach is therefore that such assisting is effectively hiding true physical and indisputable performance edges until the technology cannot cope, having the effect of 'sharpening the edges' and making it more difficult for you as the driver to be aware of the distance to the physical performance edge and to predict when you fall over. How can you ever be a truly judicious driver that continuously makes comprehensive choices about direction and speed, if technological designs elude you the edge awareness necessary for such decisions? The alternative is perhaps even worse. If technology is to have full responsibility, it must be given full authority over direction and speed, which ultimately implies to decide where to go!

7.5 Summary of difficulty characteristics

The framework, the character of uncertainty (R1.2.1), is intended to assist in analyzing and understanding why situated control may be required also with a detached and rational view, without even considering the ideological (i.e., moral) reasons for situated control introduced in chapter 6. The constitution of the world, the activities, and the systems that we use, tend to imply at least some of these uncertainties. By using the contributed framework, the idea is that it will influence the approach to controllability during design thereby assisting in achieving a beneficial balance within the tension between effectiveness and safety.

With an understanding of the necessity of situated control during design, the character of controllability as the design result might become different. The extension of the tension between effectiveness and safety to also include the concepts of potentiality and resilience assists in putting focus on the distinction between detached and involved perspectives, a distinction associated with calculative and generative design and safety cultures. The calculative culture is here connected with a focus on the detached perspective and the generative culture with the involved perspective, all these words are keywords in the character of control (R1.2.2) discussed in chapter 8 as well as in the contributed redefinition of usefulness (R2.2) provided in chapter 10.

[38] cf. footnote 18, p. 63, and the corresponding notion of a flight-envelope for aircraft.

8 Models and scenarios, and simulators

The only thing certain about a scenario is that the future will not be exactly like it, a rhetoric argument I have felt compelled to use at numerous occasions in military contexts where reliance on scenarios are commonplace. Scenarios are, of course, important and valuable tools for planning, if proper considerations are made about their shortcomings. The point of the argument is that planning and preparation is a kind of optimization that implies a deliberate reduction of preparedness for aspects not covered by the scenario. The problem is that if aspects of the scenario are taken too seriously, it might result in preparations so specific that situations actually covered become difficult to handle because of mismatching details. Another point I often find myself trying to make is that simulators can never achieve real-world experiences and therefore never replace real-world training. Real-life skills require real-life experiences and artificial skills can never be more than complementary. Sterman (2002) reflects on becoming a systems scientist and maintains like Box (1976) that all models are wrong. These three statements are essentially equivalent, addressing the same principal issue, the fact that the nature of models, their core property, is both a strength and a weakness, and scenarios as well as simulators are models.

The core purpose of a model is to clarify and explain things and in order to establish some kind of explanatory power greater than what is natural for whatever reality that is to be explained, certain simplifications and reductions of scope and complexity must therefore be performed. Otherwise, there would simply not be any benefits from producing the model. This means, however, that the resulting model by design is a reduction of richness that while increasing clarity also risk making the understanding irrelevant (cf. the scientific dilemma of rigor versus relevance, discussed in ch. 2). The necessary abstraction and generalization that is the core of every model implies also a detachment from context. In essence, a model becomes a statement that although it may be perspicuously valid yet is missing the point. To regularly rely on models being correct resembles in fact the assumption that a particular joke is always appropriate, regardless the occasion. The more severe example of model misapplication is to assume regularly that a modeled behavior, for example implemented as an autopilot, always will be safe. While the effects of these two examples of a maladaptive use of models are vastly different in character, they have, principally, a common cause, which is to mistake the map for reality.

Models, simplified descriptions, and maps, they are invaluable for understanding principles and for making complex systems comprehensible. Real systems do not work exactly as the models, and the difference may be what matters. Models are stereotypical representations of whatever is modeled. Scenarios provide stereotypical descriptions of whatever situation it describes. Simulators create stereotypical experiences according to implemented simplified representations of whatever is simulated. Stereotypical and detached understanding is valuable information for rational and analytical decisions. It is the 'ideal' input to system 2 (i.e., intelligible with a clarity that exceeds the blurry reality), but system 1 comes first, is quicker, relies on an embodied (i.e., situated) sensing of reality, and system 1 is what human beings use the most (e.g., Kahneman 2003, 2011, Klein *et al.* 2010). The theoretical framing selected for this chapter is research focusing on the human role in relation to technological systems.

8.1 The human contribution

When unpredicted things happen around a technological system, it may result in a breakdown or a close encounter, or perhaps it becomes a great success. For the latter case, which clearly is a desired result, it is common to attribute the success to clever human performances and vigilant seizure of opportunities. For the undesired case, on the other hand, the assessment is more open. Are human system operators the culprit, or heroes? When a system actually breaks down, the judgment tends to be against the human component. Causes of disasters are often assigned vaguely to the human factor, or more precisely to various forms of human error (Reason 1990). However, when there is merely a close encounter, which arguably is a situation very similar to the breakdown situation, except for the actual outcome, nevertheless a situation depending on the human factor (Vicente 2006), the conclusion is often the opposite. The same characteristic human behavior that for the undesired outcome was considered the culprit, represent a valuable human contribution (Reason 2008).

For most cases with undesired results involving both human beings and technological systems it is with hindsight (presumably always) possible to find aspects of individual human behaviors that may have exacerbated the situation, but it should likewise be possible to find aspects of technological designs that contributed to the outcome. It seems, however, to be closer at hand to blame people than to blame technology, presumably because technology is intuitively considered to lack incentives and the possibility to take responsibility.[39] How to view the situation is a matter of attitude, and the

[39]On the other hand, there is this trend of implicitly endowing technology with incentives by using concepts such as autonomy for technological systems, which implicitly implies taking responsibility (cf. ch. 1.4.1, 10.3, 15.1).

discrepancy between the viewpoints may be taken as represented by the clashing engineering and ergonomics perspectives (Kirwan 2000). One fundamental problem is that viewing things from a detached (e.g., engineering) perspective decouples events from meanings, allowing for statistically rigorous conclusions and adherent interpretations of causes (i.e., courses of events) that in practice are determined by the questions asked, like when 'proving' that doctors are more dangerous than gun owners (Dekker 2007). Before the birth of the human factors research discipline, focus was on designing the human to fit the machine. That is, the engineering perspective came first and the training of operators became a consequence. With experiences mainly from military flying where things are brought to the extreme, it became evident that the technology must explicitly be designed to fit human beings (Wickens and Hollands 2000). However, focus appears still often be on detached aspects and concrete performance measures, arguably an indication of a prevailing paradigm of technical rationality favoring the detached engineering perspective. Although, it appears there is a growing stock of research and writings focusing on the ergonomics perspective, on the situated nature of the human role, and on problems with technology essentially designed according to detached aspects. For example, the irony of automation (Bainbridge 1983), the human-centered approach (Billings 1997), resilience engineering (Hollnagel *et al.* 2006), human-tech (Vicente 2011), and other consequences of the detached perspective, by Dekker (2013) argued as originating from a Cartesian view on consciousness. The present research builds upon such writings and aims to contribute in drawing focus to the situated perspective.

The human role, and thereby the human contribution to overall system performances, have changed with the evolution of technology, in particular since computers became commonplace (more about the human role in ch. 9.1). Computers have an unprecedented capability to become involved with human performances, in a way that more mechanical machines cannot come close to, due to the logical powers and dynamic alteration of properties facilitated by microprocessors and software designs. This increasing capability to get involved can be illustrated by the notion that the computer reaches out (Grudin 1990) or by the three paradigms of HCI (Harrison *et al.* 2007). Initially the computer was viewed as a machine to be controlled, later as a thing to be worked together with, and today as a socially collaborating device. The computer reaches, so to speak, out from the nuts-and-bolts of the mechanic domain into the social world of relations and meanings. Consequently, the attitude towards computers change and new paradigms arise. At first, machines were machines and people were people, today there are notions of joint cognitive systems (Hollnagel and Woods 2005, Woods and Hollnagel 2006) and distributed situation awareness that includes the machines (Stanton *et al.* 2006, 2010). While it might exist well- grounded reasons to speak about technological awareness and artificial cognition for

example in order to design systems that resonate better with human beings, arguably there are fundamental qualitative differences between human and artificial variants of awareness, intelligence, and cognition. The present stance is that the human contribution is to be human, and in order to support that role differences between people and machines must be explicated rather than blurred, which makes the talking about joint cognitive systems and distributed situation awareness that includes artificial components problematic (elaborated on in ch. 15.1).

8.2 The character of controllability

There have already been put forth two different characters of controllability, the approach to controllability for the case of design and the character of controllability as the design result. For the case of design, the notion of different cultures was used to distinguish between attitudes towards controllability. On the one hand, a calculative culture aims to reach predetermined goals along predefined routes predicted as safe and effective, by designing systems with predictable properties and well-defined workings. On the other hand, a generative culture aims to reach contextually relevant goals along routes not yet known in comprehensive detail thereby associated with uncertainties and hazards, by designing systems with properties providing potential and resilient system workings. Such culturally dependent assumptions, explicit or subconscious, become because of decisions made in the case of design implemented as a kind of technological prejudice affecting the character of controllability as the design result. The case of use is taken as a significant summary for the overall character of controllability, a summary that connects control issues with the psychological dimension of system controlling individuals.

The character of control, the case of use
Detached ↔ Involved

The tension between a detached and an involved character of control, the character of control in the case of use (R1.2.2), summarizes the tensions within the approach to controllability, the case of design and the character of controllability, the design result. This tension within the character of control seems to be what research about levels of automation, supervisory control, out-of-the-loop performance problems, issues with decision support systems, and such like, is about. These topics will be discussed further in chapter 9.1 – The human component.

8.3 From conditions to effects, in practice

In chapter 6.3 – From conditions to effects, in theory a map was presented showing the principles of how technological and environmental conditions become effects through human activities governed by psychological aspects. The physical domain constitute the concrete conditions that by appearance and availability of controls for system controllers inflict psychological incentives (e.g., will) to apply system properties in the situation. Available controls provide also the means by which intended effects are pursued, if there are sufficient skills to control the system such that the goals are reached. Concrete physical effects occur as the result of actual control activities, which in turn result in psychological (and physiological) experience. The principal map was illustrated by Figure 6.1 (p. 117).

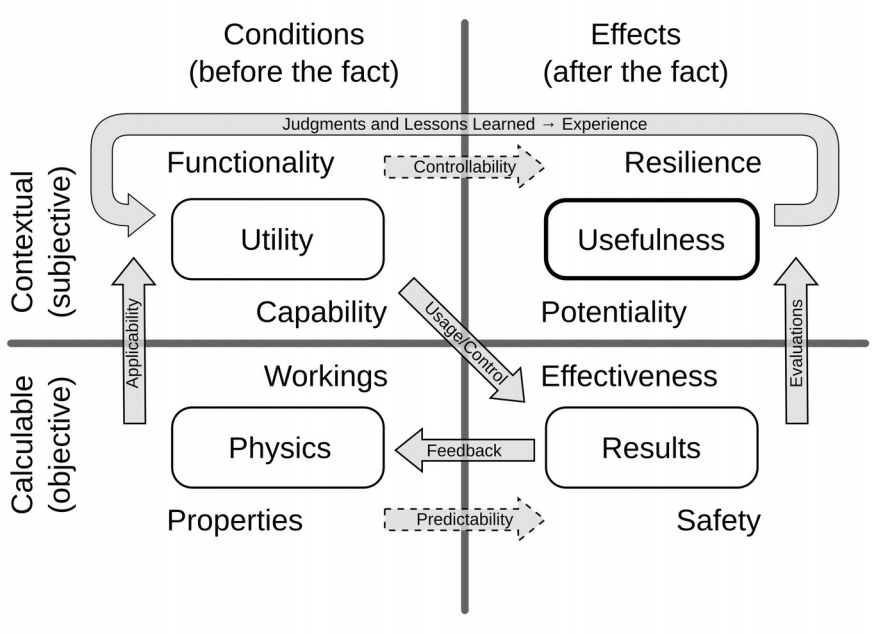


Figure 8.1: From Conditions to Effects, in practice (R1.2.3), i.e., walking the map

Figure 8.1 (R1.2.3) is essentially the same map as Figure 6.1, but with the aim to depict more fully what actually happens in practice (i.e., in a real case of use), by focusing on the character of controllability and by defining the present view of certain usefulness-related concepts encountered while pursuing the research questions. The domains are the same as for the map, conditions are before the fact and effects are after the fact. Conditions are categorized as calculable, ideally objective, concrete and physical aspects described by system properties and system workings, or as contextual conditions in the form of subjective interpretations described by concepts

such as functionality and capability (these concepts as well as system properties and system workings were explored in chapters 5.3 and 5.4). Effects are also categorized as calculative (or calculable) or contextual, a distinction explored in chapters 7.3 and 7.4. The concepts of effectiveness and safety were taken as describing calculative results while potentiality and resilience describe more contextually oriented effects.

The relation between these concepts are, presumably, best explored by scrutinizing how aspects they intend to depict in fact come to be, which is what Figure 8.1 aims to depict. The appearance of the concrete conditions, naturally shaped by biases and presuppositions coming from knowledge and experiences (or lack thereof), is the foundation for interpretations made by system controlling individuals about applicability. Depending on purposes and goals, the applicability is in turn the foundation for interpretations about functionality. Based on assessment of efficacy of system properties within the actual environmental setting (i.e., assessing whether the thing actually can do the work during present circumstances), the intended function of the technology is identified. That is, the present definition of functionality is that it denotes the set of intended functions a system might fulfill. While system properties and system workings are considered essentially overlapping concepts, the difference is that workings indicate dynamic properties (e.g., machine properties, P3, ch. 5.3) and properties denotes more static and fundamental aspects. Between functionality and capability, the relation is considered somewhat similar. Functionality refers mainly to properties potentially having a function while capability depicts functional aspects over time as the result of system workings.[40] The identified applicability of properties and workings, in other words, the assessed functionality and

[40]As a side-note, associated with the present discourse about using concepts with human connotations when talking about technological systems, 'capability' sometimes appears mixed-up with 'ability', encyclopedia definitions follow:

Capability: 1 : the quality or state of being capable; also : ability. 2 : a feature or faculty capable of development : potentiality. 3 : the facility or potential for an indicated use or deployment <the *capability* of a metal to be fused> <nuclear *capability*>. http://www.merriam-webster.com/dictionary/capability

Ability: 1 a : the quality or state of being able <*ability* of the soil to hold water>; *especially* : physical, mental, or legal power to perform. b : competence in doing : skill. 2 : natural aptitude or acquiring proficiency <children whose *abilities* warrant higher education>. http://www.merriam-webster.com/dictionary/ability

While the above definitions actually allow for non-living things to have 'ability', they still indicate a subtle difference where 'ability' is inclined towards human qualities, skills, and mental properties. Therefore, the suggestion is to be careful with 'ability' in order to save the richer aspect to human contexts. 'Capability' is perhaps already a too rich concept for technology, although it appears required for extending the slightly narrower notion of functionality. However, such usage is nevertheless a potentially counterproductive development, this problem is explicitly discussed in (ch. 15.1).

capability, in relation to the practical worth of estimated effects, constitute the utility associated with the technological system of concern (ch. 6).

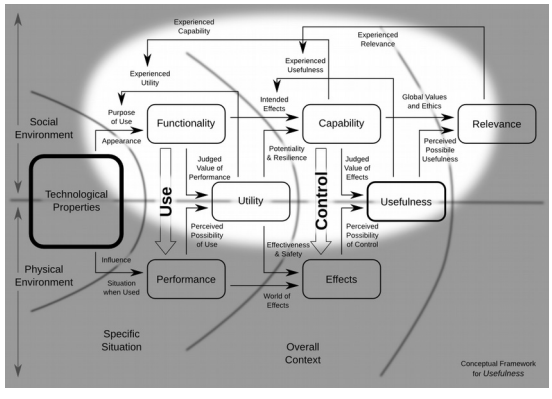
For effects to occur, however, some kind of human activity is required, implying usage or control of the technological system. The main purpose of this model is to illustrate the impact of the character of control (R1.2.2). A calculative approach focusing on predetermined system behavior resulting in a detached character of control is argued to confine effects to the lower half, depicted by the dashed predictability arrow (cf. the rising boundary of the calculative domain in Figure 6.1). While on the other hand, a generative approach focusing on contextual aspects resulting in involved control is argued to open up to effects in the upper half as well, depicted by the dashed controllability arrow. This difference is because the predictability approach largely opposes situated controllability. Whether results actually are considered as having been provided by potential and resilient system effects is yet another matter of subjective interpretations. The usefulness of system properties is ultimately determined by subjective evaluations that become incorporated in the psychological domain as experience and lessons-learned.

8.4 Summary of practice models

The character of control (R1.2.2) is essential for what effects that in fact will occur, whether they will be calculative (i.e., predictable, stereotypical, model-based) or generative (i.e., local, contextual, situated), simply because the character of control is the link between conditions and effects. Actual control within a situation, as opposed to intended control in modeled and predicted situations, is what distinguishes 'in practice' from 'according to theory' in the present context of getting from conditions to effects. For human beings the level of involvement seems in practice crucial in a number of significant ways, of which the most important perhaps is the effect involvement has on how we interpret things. Hence, the character of control is not only significant for what effects that will occur but also for how we value them.

The human contribution to system effects is to enact this link between conditions and effects, which then becomes to walk the map in practice (R1.2.3). For this walk the character of the technology used is, obviously, significant. Therefore, this model (or typology, or map) aims to keep focus on the contextual, it aims to counteract forgetting about the involved perspective as a consequence of technological structuralization forcing focus to be merely on detached aspects making the success of calculative effects a self-fulfilling prophecy. The model can be used to explain the difference between model-based and situated usefulness.

9 Situation awareness



If the premise can be accepted that an inescapable aspect of the human role is to be responsible for system effects, then consequently it is necessary to accept also the premise that human beings must have authority and be able to control systems comprehensively, such that intended effects occur and undesired effects are avoided. It is vicious to hold people responsible for effects that they cannot control, and to understand fully implications and nuances of system control people need, apparently, to experience situated effects first hand. Furthermore, to understand what is required for a human system controller to maintain authority and have the ability to control a system comprehensively, it is necessary to scrutinize the human being as a sociotechnical system component. This kind of scrutiny belongs to the rather diverse research area labeled human factors (and ergonomics), an area that reach from physiology and bodily strains in workplaces, via psychology and mental workloads, to sociology and research about effectiveness and relevance of different technologies. This diversity with its adherent problems of academic positioning is reflected in disciplines such as Human-Computer Interaction (HCI) and Information Systems (IS) research, disciplines that sometimes appear slightly lost in the academic universe (e.g., Orlikowski 1992, Orlikowski and Iacono 2001).[41] For the present purpose, mainly

[41]These references are added as examples of the debate within IS research about what actually is the object of research. For the HCI field I would simply like to refer generally to the ubiquitous debate at HCI conferences about the whereabouts of the field. Another example of the somewhat vague identity is the fact that HCI research at Uppsala University is split between one department in the social sciences faculty and one in the natural sciences faculty.

focusing on the specific case of situated use of technological systems, the human component is limited to the role of being a system controller. For scrutiny of this role, the ubiquitous human factors [in aviation] buzzword, situation awareness, is taken as a starting point. It is a concept that on the one hand is considered a purely psychological phenomenon (often taken as represented by Endsley 1995b) that arguably has come to have connotations indicating that it primarily is awareness about aspects in the environments of the controlled technological system. On the other hand, situation awareness is sometimes considered possible to be distributed among both human beings and artifacts (e.g., Stanton *et al.* 2006). The presence of these clearly incommensurable views is one reason why situation awareness is a starting point for discussing the human role. In order to function as a comprehensive and informed system controller it is, however, necessary to be aware of all aspects relevant for controlling the technological system of concern, which then also includes awareness about the technological system itself and about interaction properties between the system and its environments. For this purpose, two additional concepts are suggested, namely system awareness and edge awareness.

9.1 The human component

Human beings are complex creatures, and as a significant component in sociotechnical systems that perhaps are better described as socio-natural or socio-physical systems (ch. 5.1), it is obviously a flaw not to consider the entire human being. That is, however, an unsurmountable task requiring overall comprehension of several scientific disciplines including medicine (e.g., physiology), psychology, as well as human factors and sociology. Our bodies are always subsumed in the physical world and phenomena associated with our minds are immersed in a social environment, making our being in the world constitute a continuously morphing and circularly dependent system component interacting simultaneously both with social and natural aspects. Arguably, neither physics nor medicine, psychology, or sociology, provide on their own sufficient grounds to claim full understanding of the human being, implying that a mere thesis cannot look into more than a tiny fraction. Yet there is, in some sense, for the present purpose a need to consider the human being as a whole.

Focus is on the human component as a controller of technological systems, for which bodily performance (i.e., strength and skill) traditionally has been essential. With the industrial revolution followed, however, a reduction of the need for physical strength and accordingly a reduction of strength as a valuable human quality. The current development, sometimes spoken of as an ongoing information or information-technology revolution, shows a similar pattern, but for certain mental abilities instead of physical

ones. As the capabilities of computerized technology increase, the need for computer-like intellectual skills is reduced and the value of these particular aspects of human thinking becomes reduced accordingly. Computers excel in formal information processing and the capability that used to be an exclusive human trait is now performed with higher precision by machines. Simultaneously, for precision reasons, for predictability and model-based safety, computerized logics and automations are implemented to handle low-level control of systems, while human operators remain in supervisory positions where they still are expected to maintain overall system control by performing high-level decisions that computers not yet are capable of making with sufficient satisfaction. Arguably, but largely disregarded by the contemporary view of usefulness, there is more to human mental abilities than formal information processing.

Furthermore, the reduced value of embodied skills and the increased focus on rational high-level and intellectual performance appears thoroughly counterproductive. As long as human beings are expected to make insightful and meaningful decisions, the body must be involved, simply because our understanding of things is highly influenced by what we do, whether we like it or not. Whatever the body is part of affects the 'gut-feeling' of things and these gut-feelings are, as it seems, our primary governor of behavior and, apparently, our primary source of options for decisions, both for routine options (i.e., biases and heuristics) and for new options and novel ideas (i.e., the trigger of curiosity and basis for our natural variability). This intuitive system appears actually also be what in the end settles most of our decisions (e.g., Klein *et al.* 2010), unless we somehow feel that it is necessary to meddle in with effortful rational or analytical thinking (cf. the three levels of car driving and flying discussed in ch. 3.1).

For human beings, the intuitive system, or system 1, is fast, always influential, often pretty good, but thoroughly afflicted with biases and dependent on heuristics, while the rational analytical system, or system 2, is slow, requires deliberate and demanding efforts associated with certain limitations, but mostly capable of producing informed and autonomous decisions (Tversky and Kahneman 1974, Kahneman 2003, 2011). This idea, or model, depicting two different but thoroughly interconnected kinds of mental processes is not new and, evidently, a model having good explanatory power. It may be traced as far back as to the reformation, if not farther, to René Descartes (Renatus Cartesius) and his distinction between mind and matter. While the scientific community since the Cartesian era has rejected fundamental dualism as well as spiritualism and superstition, the Cartesian distinction has prevailed and a discourse has continued about which way to look at things. The realist-empiricist-positivist tradition argues for a matter first – mind after ordering, while the relativist-idealist-interpretivist tradition goes in for mind first – matter after (cf. the discussion about philosophy of science in ch. 2.1). One consequence of this debate is that the mind has

somehow been decoupled from the material body, a disconnection facilitating, and perhaps also exaggerating, the purely rational and logical view of the mind. It appears at least that models describing the workings of the human mind, or the human being in general, in not perfectly rational terms (i.e., scientifically proven according to the present paradigm) tend to be much disputed. Perhaps are many of these objections primarily due to the inherent difficulty in describing and studying psychological phenomena not directly observable or not observable in isolation?

The information-processing model "provided a new way to study the mind" (Goldstein 2007, p. 13) and, for example, the notion of long-term memory and short-term working memory associated with a limited number of (e.g., 7±2) input and output channels (Miller 1955) make certain human characteristics understandable. Along with the advent of computers, the information processing approach to studying the brain transformed into a computer metaphor of the brain. However, metaphors feed on superficial similarities and they work both ways. It is therefore possible to ask whether the strength of the computer metaphor of the brain comes from helping us understand the brain by viewing it as a well-known system, or if it comes from helping us understand the computer by considering it similar to a brain that we understand intuitively? The problem is that the metaphorical comparison obscures fundamental differences. Arguably, the computer metaphor and the information-processing model are most relevant for studying and describing the rational and analytical system, system 2. Using it as a model for human intellectual performance in general, risk thereby obscure aspects related to system 1 and aspects related to interactions between the two systems. The murky shadow of a powerful model is the risk that it becomes considered 'the truth'. Without alternative models with comparable explanatory powers about other aspects there is a risk for these obscured aspects to become forgotten, a tendency described by Kahneman (2011) as 'what you see is all there is' (WYSIATI). The explicitness and describable character of system 2 aspects, and thereby the enhanced possibility of studying these aspects in isolation, work in favor for the computer metaphor of the brain, a development that unfortunately appears to work against the nature of human beings.

The dual model of the mind does therefore provide useful explanatory powers for the present purpose. Hence, human thinking is here described as essentially having two modes of operation, fundamentally different in character, but not independent. On the one hand, there is the focused, controlled, voluntary, rational, analytical and arithmetic system 2. On the other hand, there is the automatic, involuntary, intuitive and emotional system 1 (Tversky and Kahneman 1974, Shiffrin and Schneider 1977a, 1977b, Stanovich and West 2000, Birnboim 2003b, Kahneman 2011), where particularly the intuitive and automatic system 1 appears thoroughly interwoven with the physical body. The model recapitulated in Figure 9.1

with overlapping fields of process and content, is considered to illustrate rather well the interconnectedness between the two systems. What you know explicitly (content) depend on how you get to know it (process) and the way you get to know things (how) depend on what you know (what you choose to do), a continuously ongoing parallel and mutually dependent (circular) loop of life, mirrored also in the map of how conditions become effects in theory (Figure 6.1), by the connections between the psychological domain and the activity domain.

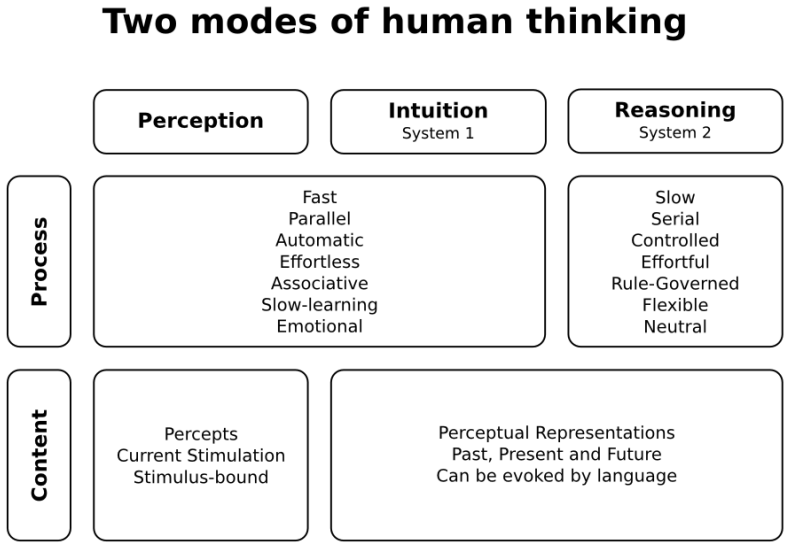


Figure 9.1: Two modes of human thinking, adopted from (Kahneman 2003, p. 698)

The interconnectedness between the two modes of human thinking is considered a sufficient argument for considering the involved perspective (Dreyfus 1972, 1992, Dreyfus and Dreyfus 1988) whenever scrutinizing the human role. For example, when building technological systems that are to be controlled by human beings they must be designed explicitly to involve the whole operator, not only system 2. Human beings make decisions as embodied entities, not as detached logical machines, a fact that makes things complicated (Kahneman and Klein 2009). In short, linear models of human decisions, often associated with the information-processing model of the mind, are here considered insufficient, as they appear to miss essential situated aspects. For example, the otherwise appealing decision ladder model developed by Jens Rasmussen while studying process industry workers (e.g., Vicente 1999, p. 187). It may be considered to lack a number of feedback connections from the 'decision and implementation leg' to the 'perception

and comprehension leg' of the ladder, feedback connections associated with system 1½ and situated interconnectedness between the perhaps slightly too clear-cut notions of system 1 and 2 (ch. 3.3, 9.3 and 15.3). Similar ideas are expressed, for example, by approaches such as dynamic decision-making (Brehmer 1992) and naturalistic decision-making (Klein 2008).

Despite the so-called shortcomings of human rationality, undesired influences of animal instincts, emotional biases and personal interests, the human role in relation to machines is to be a human being. Anything else would obviously be a mistake. To be a human being includes, however, having emotions and animal kind of instincts and intuitions, which in the end are what determine the values of effects. Usefulness is ultimately defined by human values that are interpreted differently depending on the character of involvement for the human interpreter. For insightful interpretations and relevant judgments of human values that are required for the ability to take responsibility properly, it is crucial to have sufficiently involved human system controllers. Detached roles result in stereotypical values and reduce the responsibility actually taken, to be only of the formal aspects that in some sense might be possible to lay on technology. The point is that such formal aspects are merely the explicit part of responsibility. For the ability to be truly trustworthy, tacit and involved aspects of responsibility must also be present, which require the tacit kind of insights that comes from personal involvement. Such involvement is what requires systems to be explicitly designed for situated human control, a requirement that stands somewhat in opposition with the declared contemporary view of usefulness focusing on predetermined effects, thereby tending to remove human control because it ruins the predictability. The alternative would be to adopt a specifically outspoken aim to design systems that allow people to become experienced experts (Dreyfus and Dreyfus 1980, Schön 1983, Dreyfus 1986, Reed 1996) that eventually also might develop practical wisdom (the present version of Aristotelian phronesis described in ch. 2.1.4).

In order to develop such human-centered or human-oriented (Billings 1997, Vicente 2011) technologies, it is assumed useful to be able to describe intelligibly what human system controllers require. For such descriptions, the concept of situation awareness is considered an intuitively relevant and therefore appropriate overarching phenomenon to start with. It is relevant for both how awareness is achieved (process), and for what it is about (content).

9.2 Situation awareness and system awareness

Situation awareness may to begin with be described as a common-sense concept for some kind of desirable human trait. It is intuitively true that people (and, if considered applicable, automatically acting machines) having good situation awareness, know what to do and are therefore not easily

fooled, thus clearly implying a desirable thing to have. Having situation awareness essentially depicts 'being on top of things', a description that, arguably, have a connotation of rational superiority, which perhaps is what makes the concept align towards computable aspects. The concept of situation awareness has been in frequent use, particularly in military aviation contexts, since the First World War (Endsley 1995a, 1995b, Wickens 2008). However, by some it is considered a critical but ill-defined concept (Fracker 1990, e.g., Sarter and Woods 1991). The very nature of situation awareness is in fact quite unclear. Is it a purely psychological phenomenon confined within a single individual, is it located in the world, or can situation awareness be distributed among a group of individuals and, perhaps, also among technological artifacts? The different views have been labeled the psychological, the engineering, and the systems ergonomics approach to situation awareness (Stanton *et al.* 2010). One question is, if situation awareness describes a psychological phenomenon, does it in fact describe anything relevant or is it merely an aggregate of other, already established, psychological phenomena (Flach 1995)? Another question is, if it actually is relevant, how do we measure it (e.g., Fracker 1990, Endsley 1995a, Stanton *et al.* 2010)? Finally, if it is both relevant and measurable, what can we to do with it? Is good situation awareness a relevant indicator for successful system operation and bad situation awareness an acceptable cause of accidents, or are such conclusions merely applications of folk models (Dekker and Hollnagel 2004)?

The present stance is that the intuitive appeal of situation awareness as a concept makes it relevant. It depicts something that is intuitively important to describe, and discarding it would thereby risk having relevant aspects overlooked. However, it may be a too complex and overarching phenomenon to be used in any other way than as a frame of reference to avoid making the mistake to overlook essential aspects. Situation awareness is here viewed as leaning towards the relevance end in relation to the scientific dilemma of rigor versus relevance, thus providing meaning but lacking the specifics required for rigorous measurements and predictions of implications. The paradox seems to be that the more efforts that are made to specify the concept the less meaningful it becomes. Therefore, situation awareness and the additional awareness concepts presented below are intended mainly as frames of reference, facilitating the finding of relevant and meaningful aspects to scrutinize further. If used in this way, these concepts are what provide meaning to the measures and not what get meaning from measures. First, however, what is situation awareness?

According to Fracker (1990, p. 1), a USAF[42] Major, “Situation awareness (SA) refers to military operators' knowledge of the immediate tactical situation … among the most important subjects to be addressed by military

[42]USAF: United States Air Force

psychologists in recent years". What, however, has become one of the most frequently used definitions of situation awareness is the one provided by Mica Endsley (Endsley 1995a, 1995b, Endsley and Kaber 1999, Wickens 2008). Her definition is a follows:

> Situation awareness is 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 (Endsley 1995b, pp. 36–37).

- Level 1: Perception of the elements in the environment
- Level 2: Comprehension of the current situation
- Level 3: Projection of future status

The above definition, often taken as the typical psychological approach to situation awareness, is a definition that indeed connects situation awareness with several non-trivial psychological concepts (i.e., perception, comprehension, and projection), while also providing a slight indication of what situation awareness is about. If the concept is to be used as an umbrella phenomenon guiding what to explore further, the question what it is about (content) is thoroughly important because an idea of content (what to be aware of) must come before scrutinizing how to become aware of this content.[43] The definition of situation awareness above says that it is perception of the elements in the environment, thereby drawing focus towards system entities other than the individual having the awareness. What about system elements not located in the environment, are they important? Must not the person having situation awareness also be aware of the controlled technological system of concern? Does the system of concern reside in the environment to the individual or does environment refer to what is outside the human operator as well as the controlled system (e.g., the road and weather conditions when driving a car)? Endsley (1995b) does mention awareness about properties of the controlled technological system (e.g., levels and modes of automation), but the overall military aeronautical framing tend to draw attention towards awareness of other elements such as other aircraft, the terrain, and threatening weapon systems. To increase the clarity about what situation awareness is about, the concept is here complemented by system awareness. Both these labels are unfortunately somewhat ambiguous because the system may be an essential part when describing the situation and situation may include the system, but in a sense this ambiguity is for good because it enforces extra care when defining what is meant by using them. Having two overlapping concepts introduces a tension that require an explicit definition of where aspects belong, which supposedly lead to beneficial scrutiny similarly as for the tensions within the character of utility (ch. 6.2) and the character of controllability (ch. 8.2).

[43]This statement aligns also with the critical realist stance that ontology comes before epistemology. There must be an idea about the world before we can discuss how we can know about the world (ch. 2.1).

Situation awareness is here considered explicitly omitting awareness about properties of the controlled technological system of concern, despite that these properties obviously are included in the situation, and it is thereby defined as being about the environment of the user together with the operated technological system. *System awareness* is accordingly defined as explicitly depicting awareness about properties of the controlled technological system of concern, despite that the system could mean both technology and environment, thereby omitting awareness about the environment system. Relevant properties of the technological system, parts and structures to be aware of are what the frameworks in chapters 5.3 and 5.4 aims to describe. The above definition and levels of situation awareness may then be translated to apply to system awareness as well:

> System awareness is the perception of parts and structure of the controlled technological system of concern, the perception of physical, functional and machine properties, the comprehension of their meaning, and the projection of their status in the near future.

- Level 1: Perception of parts and structure
- Level 2: Comprehension of functional and machine properties
- Level 3: Projection of future status of system properties

9.3 Edge awareness

System awareness and situation awareness is, however, not enough. As already have been introduced (in the thesis introduction, ch. 1, and in the introduction of this chapter, ch. 9), what is often the most important kind of system effects is what comes from interactions between the technological system of concern and the environment (e.g., situated effects). Furthermore, especially level 3 of both situation awareness and of the newly introduced concept of system awareness are associated with certain problems. Firstly, projection of future status (level 3) may be difficult to distinguish from comprehension of system properties and situation aspects (level 2). Projection is perhaps an essential aspect of comprehension, thereby making the distinction between awareness of level 2 and level 3 irrelevant. Secondly, level 3 and the projection of future status is what appears to lead analysts into attributing causal powers to the concept of situation awareness (and from now on possibly also to system awareness), such as when stating that an accident was caused by loss (or lack) of situation awareness. While, in fact, the outcome from a situation always is the result of decisions and actions that, presumably, but not necessarily, are affected by situation and system awareness. Used as a causal agent, situation (and system) awareness becomes therefore a circular argument because decisions and their effects are caused by the phenomenon on which they are based (Flach 1995).

The concepts themselves, however, are here considered highly relevant, while the levels are merely additional aspects possibly of enriching and illuminating character. Level 3, the most problematic notion, ties in with application and situated control, which may reach across domains and connect the state of being aware with the process of becoming aware and making decisions based on that awareness. This intertwined domain is, however, what edge awareness is about. For the present purpose it is not necessary to sort out whether the concepts are psychological phenomena or part of psychological processes, it suffices to recognize the situated and applied character of level three, to which edge awareness belongs. Edge awareness lies in the borderland between system 1 and system 2, it is thereby not possible to categorize as either rational or intuitive because it is both. Hence, the notion of system 1½ suggested in chapter 3.3 (and ch. 9.1). Edge awareness is a concept of a situated phenomenon thereby not possible to categorize as either process or state, it is impossible to speak about what it is without considering how to get it. Edge awareness is, arguably, a concept closer to what the early aviators meant by situation awareness (e.g., Fracker 1990), closer to the practical notion of 'being on top of things', and a concept to be contrasted against the more theoretical psychological state described by Endsley (1995b) or against the distributed phenomenon that may include technology (e.g., Stanton *et al.* 2010). To distinguish further between the three concepts of awareness we must return to what they are about, and we have already discussed what situation and system awareness is about. So, what is *edge awareness*, and what is it about?

Edge awareness is, obviously, awareness about edges. It may be awareness about edges in the situation (i.e., in the environment), which for physical aspects often take the form of hazards (e.g., falling over a physical edge leads most often to some kind of physical damage), but also of opportunities. The most straightforward example of an environmental edge in this thesis context is, of course, the mountain plateau edge (ch. 1.4.3), an edge that will kill you if you fall over, but there are also social edges such as when passing the limit where you lose trustworthiness in the eyes of colleagues or friends. For system properties the most straightforward kind of edges are probably system modes of operation, where system edge awareness means to be aware of, for example, when the traction control system of your car engage or when the automatic gearbox will shift gears. However, the probably most interesting kind of edges is those that occur in the interactions between system properties and the environment because that is commonly where desired and undesired effects of system usage occur. For vehicles, such interaction effects are at the core of their purpose and the edges have tangible consequences. It is therefore crucial for a vehicle-controlling operator to be aware of these edges. For cars, there is for example the edge in the friction curve, and for aircraft, the corresponding edge is in the lift coefficient curve (Figure 9.2).

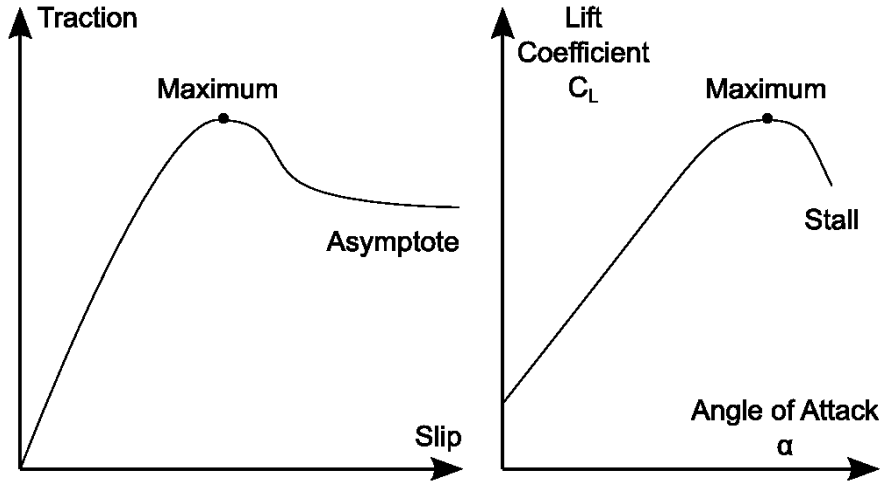


Figure 9.2: Typical friction and lift curves, examples of continuous performance relations with (intuitively) observable extremes

These two interaction relationships are examples of physical relations, objective and indisputable as far as current science knows, and always central to the maneuvering of these two kinds of vehicles, no matter the development of so-called supporting technologies. For cars, the relation between traction and slipping forces depend on friction between the tires and the surface the car runs on, and for aircraft, the relation between lift and angle-of-attack[44] of the wings depend on aerodynamics. Both these relations show clearly identifiable extremes of maximum performance, extremes that also imply a kind of instability that makes it possible to fall over the edge, for example if the driver or pilot is pushing things too far. If you skid slightly more when at maximum, the lesser friction when slipping over the edge will make you skid even more, and if trying for more lift by increasing the angle-of-attack when already turning optimally with your aircraft you will get less lift because the wings will start stalling. Another characteristic of these interaction curves is that they are continuous and often smooth, as most natural (i.e., physical) relations are, thus possible for most people to understand intuitively (i.e., 'get the hang of'). Admittedly, however, it is a kind of understanding that may take a lot of practice to master, which is what rally drivers and aerobatics pilots (and military pilots in dogfight training) do, they practice a lot to become experts in balancing on these edges with the systems they use. In the end, these performance relations are what matters, whatever the controlling system is.

Now consider again the case of traction control and computerized flight control systems. These systems may properly optimize low-level vehicle

[44]The notion 'angle of attack' was mentioned briefly in ch. 1.4.4 (footnote 10, p. 25) and is further discussed in ch. 12.1.1 (footnote 57, p. 181).

control according to such mostly well-known and indisputable physical interaction relations, and for unstable airframes, computerized flight control systems are required because the dynamics are much too quick for human beings to cope. However, what about the impact on operator edge awareness? If the traction control system imperceptibly takes over low-level control when closing in on maximum traction, thereby preventing you from slipping over the edge, which is good for statistical safety, what impact does it have on traction edge awareness for you as the driver. Traction edge awareness is arguably essential for safety-crucial medium-level decisions concerning local and situated speed and track choices (e.g., curve taking, cf. ch. 3.3), as well as for insightful high-level decisions about such things as general speed profile and overall traffic behavior. The key aspects are the words 'imperceptibly' and 'take over' because human decisions are largely governed by system 1 that is tutored by experiences that comes from perceiving and doing things first hand. The problem is that the crucial interaction relations necessary to perceive and, arguably, experiment slightly with in order to 'get the hang of' how the work, depend today not only on natural and smoothly continuous physical relations (i.e., observable and understandable),[45] they depend also on technological properties that for computerized systems may be discrete as in immediate and covert (ch. 5.4). The overall behavior of the car is significantly different when the traction control system is engaged compared to when it is not and it is therefore of crucial importance to know about these system properties that interfere with the experiencing of such a fundamental performance relation as the friction-traction edge. The problem with most computerized systems is that they are often rather bad at providing intuitively understandable feedback about when they are about to shift modes or break down.

Arguably, the present approach, the contemporary view of usefulness, is to focus on the calculative (e.g., statistical) result, which is to keep vehicle systems in general on the right side of the edge. To make vehicles in general stay on the right side of the edge is obviously a desired result but still a genuinely detached approach. From the detached perspective it suffices to be, or to have been, on the right side of the edge, a matter predominantly evaluated after-the-fact thus detached in time, thereby gaining the character of a measurable state thus detached in space and time as well as emotionally detached (cf. ch. 7.2). What is forgotten is how to stay on the right side of the edge, or how to get back if circumstances somehow have made someone ending-up slightly over the edge, which is a matter of managing a dynamic system in real-time. The how is not only an inherently involved matter, it is also a fundamentally overarching matter because of its impact on high-level decisions. The low-level control performed by the traction control system

[45]The specific physical interactions are perhaps not directly observable, but immediate effects on the vehicle as a whole can be sensed, i.e., motion patterns, vibrations, sounds, etc…

does not help very much if it has made the driver enough unaware of the traction edge to make the high-level decisions to skip putting on winter tires and keep summer tempo on the highway under snowy conditions. When the detached approach dominates and the involved perspective is forgotten, the result tends to become enforcing solutions and systems telling the user what to do. The same technological capabilities, however, should with another approach to the human role, be able to help the user understand what to do, which then, arguably, is to make the user understand how to do it. The telling system have the implicit effect of hiding the crucial edge while a truly helpful system should enhance the edge and improve the edge awareness.

Telling systems tell about predetermined truths. At the other end of the spectrum of predetermination, lies the aspect of seizing opportunities, which actually may be another key factor for non-calculative safety. When a vigilant and proficient system controller intuitively saves a potentially hazardous situation by following a hunch and engage in a behavior that lies outside the normal routine (e.g., Reason 2008, or 'The miracle on the Hudson', National Transportation Safety Board 2010), the probability for success is highly dependent on knowing system performance limits. Edge awareness is thereby crucial both for avoiding to fall over edges (i.e., for safety), and for knowing how far to push the limits in order to seize a rare opportunity (i.e., for potentiality). Without it, overall system behavior will remain with the main predictable stream, an ideal situation for adversaries as it scatters their fog of war, or simply a system bound to fail as the result of being a tightly coupled system of high complexity (Perrow 1999).

Having situation awareness depicts, in the present context when complemented by the concepts of system awareness and edge awareness, the quality of being in control regarding environment conditions. Situation awareness includes knowing about relevant aspects in the environment, a kind of knowing-that things are what they are (level 1). Having situation awareness depicts also the quality of knowing what to do about it, a kind of knowing-how to make use of level 1 situation awareness (i.e., level 2 and level 3). Especially, it means to be aware of the situation in terms of an ability to handle hazards and make use of opportunities.

Having system awareness depicts accordingly the quality of being in control regarding physical, functional, and machine properties of the controlled technological system of concern. It includes knowing about system structure and workings, knowing-that system properties can achieve certain effects (level 1). Having system awareness depicts also the quality of knowing-how to handle these properties such that effects become what are intended (level 2 and level 3). Particularly important is therefore to know about and know how to handle functional edges affecting system performance, and thereby know about what effects that probably will result from system usage. Mode shifts are typical functional edges.

Having edge awareness is slightly different. It could perhaps be described by using the same three levels as for situation and system awareness, but essentially it depicts an applied or situated level 3, a combination of overall situation and system awareness as well as awareness about dynamic interactions between the domains that the first two awarenesses are about. Edge awareness is not primarily about a distinct domain of its own but about situated aspects of the other domains in conjunction. In particular, edge awareness is about the performance edges that occur between system and environment properties. The three awareness concepts are summarized in Figure 9.3 (R1.2.4). Arguably, edge awareness is to situation and system awareness what phronesis (practical wisdom) is to techne (scientific knowledge), it is a practical or applied situation and system awareness.

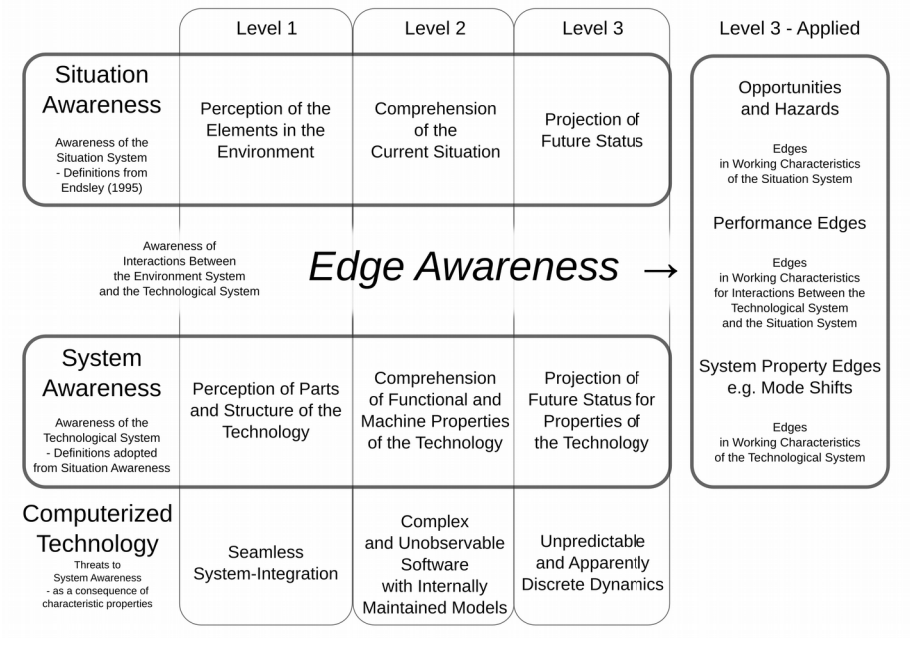


Figure 9.3: Situation Awareness, System Awareness, and Edge Awareness (R1.2.4)

The characteristic properties of computerized technological systems are directly relevant for the possibility of maintaining appropriate system awareness, thereby having impact on possible edge awareness. The listed properties (ch. 5.4, C1 – C3) may be interpreted as affecting the levels of system awareness in reverse order. Seamless system integration (C3) makes it difficult to know about and understand system structure, complex and covert software models (C2) makes it difficult to understand the implications of system properties, and unobservable and discrete dynamics (C1) makes it difficult to maintain awareness about functional edges.

9.4 Summary of psychological concepts

The human role is to make sense of things and take (personal) responsibility, which is a situated activity by the human component, not possible to reduce to rational decisions based on detached measurements. “Sensemaking is about the interplay of action and interpretation rather than the influence of evaluation on choice. When action is the central focus, interpretation, not choice, is the core phenomenon …” (Weick *et al.* 2005, p. 409). Edge awareness (R1.2.4), together with situation and system awareness, seems thereby closely related to sensemaking as both are about situated and dynamic interpretation and action (system 1½). The rich characteristics of meanings and interpretations, as well as the nuances of skillful activities, appear all largely be of tacit nature, problematic for scientific scrutiny. “The language of sensemaking captures the realities of agency, flow, equivocality, transience, reaccomplishment, unfolding, and emergence, realities that are often obscured by the language of variables, nouns, and structures” (Weick *et al.* 2005, p. 410). Therefore, the awareness concepts presented here are intended as guiding concepts, concepts possibly providing meaning to detached measures by constituting a relevant frame of reference. Simply put, it makes sense to have good edge awareness.

10 Explaining usefulness

10.1 Conceptual framework for usefulness

Throughout the previous chapters of phase one, several important concepts related to usefulness have been explored, concepts such as utility and relevance, functionality and capability, performance and effects, use and control, and a few additional ones. Several models, typologies, and conceptual frameworks have been presented, connecting these concepts, describing their aspects, and involving even more concepts. The framework (R2.1) illustrated by Figure 10.1 is an effort to bring it all together.

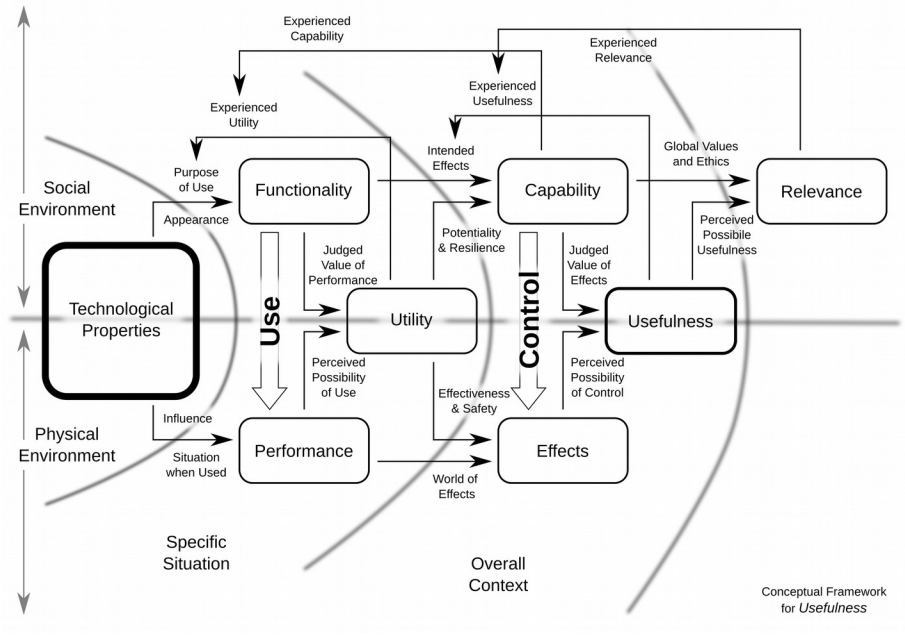


Figure 10.1: Conceptual framework for usefulness (R2.1)

The framework above (Figure 10.1), aligns with the sociotechnical (socio-physical, socio-natural) outline presented in chapter 5.2 (R1.1.1) distinguishing between the social and physical environments (the upper and the lower halves in this figure), and that makes explicit the technological system of concern (the highlighted rounded box). The worldview (R1.1.1)

corresponds essentially to the first field from the left (delimited by the first arc). Chapter 5.3 and 5.4 (R1.1.2 & R1.1.3) explores the nature of (computerized) technological properties (P1-P3 & C1-C3).

Within a specific situation (the second arc from the left) the objective aspect of usefulness concerns what influence (arrow from technological properties to performance) the properties of the technological system can have on the situation when [they are] used (the physical aspect of a specific situation). The subjective aspect of usefulness in a specific situation is how system appearance (arrow from technological properties to functionality) influences the purpose of use (a socially situated aspect emergent in a specific situation). Functionality is defined as the set of possible functions the technological system might fulfill, functions that in the end depend on the purpose of use and on the situation at hand. What functions the technological system actually will fulfill, its performance, depend ultimately on how the system in fact is used (the thick arrow from functionality to performance). The utility of the technological system (R1.1.4) is then depicted as dependent on the perceived possibility of use and on the judged value of [its] performance, a concept in which there is an inherent tension between achieving desired effects and avoiding undesired effects (ch. 6.2). Chapter 6.3 provided a map (R1.1.5), showing how properties and situational conditions become effects, in theory.

Within the overall context (the third arc from the left), the physical aspect of usefulness is what effects the performance of the system creates. The capability of the technological system is depicted as a subjective judgment (a social aspect) about how the perceived functionality can achieve intended effects. The difference in the character of uncertainty (R1.2.1, ch. 7.2) between the physical world of effects and the social world of intended effects explains in part the tension within utility. While calculative predictability (often appropriate for essentially physical aspects) sometimes can achieve effectiveness and safety, a generative approach to controllability seems required for potentiality and resilience (ch. 7.3 & 7.4), as well as for contextual relevance. The character of control (R1.2.2, ch. 8.2) as the result of the approach to controllability is influential on both skillful (physical, hands-on) low-level use and on rational (supervisory) high-level control of the system. What effects that in the end occur, depend on actual application, which inescapably is a human activity illustrated by the map (R1.2.3, ch. 8.3) showing how conditions become effects in practice.

Situation awareness was explored in chapter 9.2, complemented with a corresponding concept labeled system awareness referring to awareness about system properties (relevant for perceived possibility of use). These two concepts of awareness was further complemented by the concept of edge awareness (R1.2.4, ch. 9.3), explicitly referring to awareness about edges in system property characteristics, in environment characteristics, and in performance characteristics for interactions between system properties and

environments. These three awareness concepts are all relevant for the concept of usefulness as it is depicted by this framework because they are fundamental to situated use and control.

The conceptual framework in Figure 10.1 (R2.1) can also be used to illustrate the essential difference for the characteristic recursiveness of systems thinking between the lower half, depicting the physical environment, and the upper half, describing the social environment. Physical properties are rather easily incorporated in recursive systems of systems models, while social systems are more difficult to decompose into recursive system components. The physical world was labeled mediocristan by Taleb (2010), described as a self-regulating and rather predictable country, while the social world was named extremistan, characterized as a self-enforcing and thoroughly unpredictable country. The latter characteristic is illustrated and possibly explained by the feedback loops in the upper half of Figure 10.1 forming circular dependencies between several socially dependent aspects (e.g., experienced utility, capability, usefulness, and relevance). Because of these characteristics and referring to the different strands of systems thinking (ch. 5.1 & 6.1), the technological system (the physical aspects) and to some extent certain aspects of the lower half (the physical environment) appear rather susceptible to analytic reduction. A reduction facilitating hard systems thinking and recursive bottom-up (from left to right in this framework) descriptions with rather predictable effects (i.e., materials → artifacts → machines → performance → [physical] effects). While, on the other hand, aspects of the social environment seem better described top-down (right to left) by holistic or soft systems thinking, descriptions that eventually may boil down to rather specific design requirements (i.e., relevance, usefulness, and capabilities [overall context, the world] → utility and functionality [specific situation] → properties [machines]). Neither approach is sufficient, nor without ambiguities. These different characteristics are here interpreted as supporting the critical realist view of stratification and emergence. The top-down view is required for understanding emergent social phenomena, while the bottom-up view facilitates analysis of mechanisms at lower strata.

10.2 The character of usefulness

The contributed redefinition of usefulness for technological systems cannot be described simpler than as a distribution of several qualitative aspects. It is a rich characteristic here chosen as represented by two tensions. On the one hand, there is the subjective tension (purpose-dependent and ultimately associated with ethical dilemmas) within the character of utility (ch. 6.2), about what effects the system has the potential to achieve, governed by user intentions about what to achieve. On the other hand, there is the practical

tension, skill-dependent and associated with practical and psychological dilemmas, within the character of controllability (ch. 8.2), concerning how to achieve such intended and potentially achievable effects (Figure 10.2). Situated usefulness is in critical realist terms an emergent social phenomenon, and the two tensions represent underlying structures and mechanisms explaining this phenomenon.

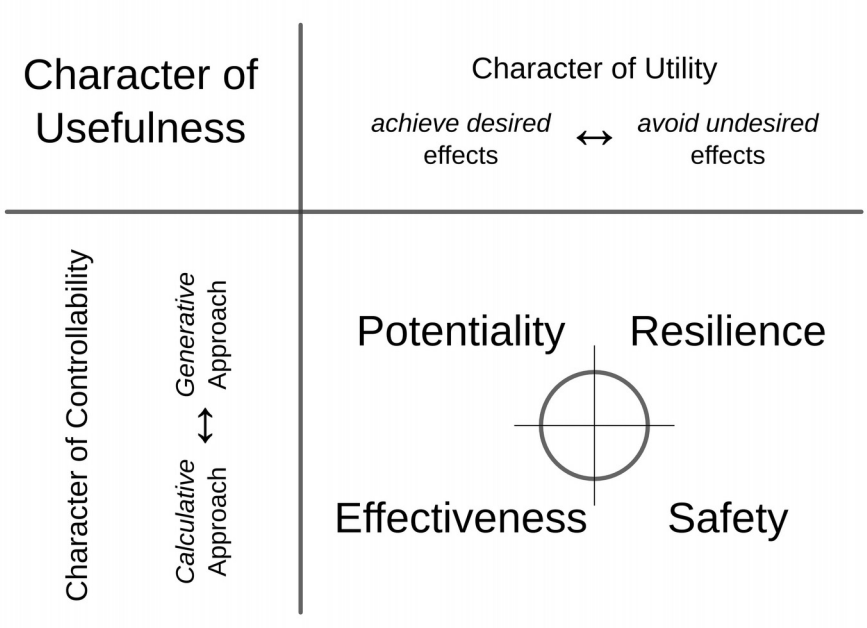


Figure 10.2: The character of usefulness (R2.2), with two tensions as its dimensions

In a sense this model (R2.2) aims to do for usefulness what Charles Perrow's (1999, p. 97) 2x2 matrix did for accidents, which was to bring together several complex mechanisms in a "tidy compression" (Weick 2004, p. 29). Perrow's model connects complexity (what) and coupling (how) with undesired outcomes (i.e., accidents). The present model connects instead aspects and mechanisms that imply complexity and coupling with a desired outcome, usefulness. High complexity and tight coupling was what Perrow described as making accidents normal, meaning that mechanisms of these kinds in conjunction bring accidents about as a normal consequence of the constitution of the system. The present model focus on what the complexity is about and on reasons for different coupling characteristics. Automation and predetermined system behaviors imply tight couplings in Perrowian terms, meaning that the more calculative the approach, the tighter becomes the coupling. What in this model thereby most closely resembles the notion of normal accidents for complex and tightly coupled systems is normal uselessness for calculative system designs associated with enough

complexity to imply ambiguous effects. If effects are non-deterministic because of uncertainties and unpredictability for both physical conditions and purposes, and the approach to controllability is too calculative, the result is a tightly coupled and complex system that is bound to be useless (i.e., of insufficient or inadequate situated usefulness) as a normal consequence of the constitution of the system. The most plausible reason for normal uselessness and accidents is, presumably, the law of requisite variety (Ashby 1956). When system control is abstracted or handed over to automation, which is a model-based abstraction, the controlling system is confined to the stereotypical, simplified, and incomplete model governing the abstraction, a model of lower variety than the controlled reality thus incapable of controlling it.

Depending on what the complexity is about, depending on the character of utility, a tightly coupled system is bound either to become useless because of brittle safety or to create stereotypical effectiveness. Systems supposed to be safe (i.e., having a character of utility focusing on avoiding undesired effects) that are bound to fail due to brittleness provide *false safety*, and systems supposed to be effective (i.e., focusing on achieving desired effects) that are bound to be useless because effects are stereotypical and thereby irrelevant provide *illusory effectiveness*. In other words, normal accidents and normal uselessness comes from a lack of resilience and potentiality.

The model (Figure 10.2) is supposed to be interpreted as follows. Because of the dimensions being tensions, the cross-hair symbol cannot be viewed as a pinpoint characteristic. It symbolizes more closely a focus of design efforts resulting in a 'hit-distribution' kind of system characteristics. Most probably, these design efforts vary for different aspects of a complex design and for the final system of systems (i.e., the designed technological system of concern), these efforts result in something that perhaps is best described as a distribution surface. If the model had been presented as a three-dimensional image, this surface would rise up from the figure like a hilly countryside field. If the model is to be used as a description of the usefulness characteristics of a specific system, such a surface is suggested as more appropriate than a cross-hair location. Although, the surface probably would have to be animated over purposes and different usage conditions.

However, the surface metaphor might still be illustrative, an assumption on which the following additional metaphors rely. To begin with, the volume under the usefulness characteristics surface is presumably bounded somehow, if for no other reason just because of limited development resources. Depicted by the tension within the character of utility with aspects counteracting each other, maximizing the usefulness in all aspects would likely require, if it at all is possible, the incorporation of disruptive technologies, which are groundbreaking new technologies capable of solving

some previously counteracting aspects.[46] Such state-of-the-art technologies tend, however, to be associated with a virtually unlimited spending of resources. Tensions, connections, dependencies, interactions and counteractions between aspects, and the bounded volume beneath it, tend therefore to make the volume and its surface viscous. If there is a peak somewhere, then there is either a dip somewhere else or a lowering of the base level (the bounded volume character). Moreover, the more distinct a peak is, the more significant are the effects on other aspects (the viscous surface character). That is, if there is no disruptive technology involved, nor several magnitudes more funding spent, it is likely that a significant peak somewhere implies a significant dip somewhere else. If it is not obvious where the dip is, the metaphor explains why it is more likely that the dip is not yet discovered than missing.

Overall, the bounded volume and viscous surface metaphors may illustrate the two-dimensional tension within the character of usefulness. In chapters 6.2 and 7.3, the horizontal tension was discussed. Too much focus on safety tends to imply reduced effectiveness and vice versa. In chapters 7.4, 8.2 and 8.3, the vertical tension was scrutinized. Too much focus on calculative effectiveness and safety tend to imply reduced potentiality and resilience, and perhaps vice versa. The last 'perhaps vice versa' indicates a need for further discussing whether the calculative and generative approaches in fact oppose each other, or whether the generative approach builds upon the calculative

It is possible to view calculative effectiveness and safety as necessary (but, arguably, not sufficient) aspects of usefulness. Without first taking hard mostly physical and calculable facts into account, without considering predictable conditions required to be met and necessarily avoided for desired effects to occur with acceptable consequences, are not generative aspects largely irrelevant? With such a view, the calculative approach may be imperfect but still provide necessary and basic, good-enough, or need-to-have, usefulness, while the generative approach merely means to add optional, luxury, or nice-to-have, aspects. However, the tension within the character of utility implying a subjective ambiguity about the purpose of use is here argued to refute this hierarchal view of the character of controllability. The hierarchal view presupposes that purposes are relevant when described formally from a detached perspective (i.e., before the fact),

[46]The invention of a 'gravity inhibitor' would definitely be an example of a disruptive technology, although that is (presumably) a science fiction fantasy. However, as the science fiction author Arthur C. Clarke says, “Any sufficiently advanced technology is indistinguishable from magic”, it is foolish to state that certain things cannot be done. Certain things computers are capable to do today would certainly have been magical yesterday. An example of what perhaps is more likely someday to become a disruptive technology is the invention of a virtually lossless energy storage technology. The invention of 'the ideal battery' is predicted to change the whole world economy and thereby alter the worlds strategic relations (e.g., Global Strategic Trends out to 2045 - Publications - GOV.UK 2014).

thereby disregarding the fact that situated usefulness is judged according to local or locally adjusted purposes. For trivial technologies with purposes that in fact are rather appropriately described by detached means, the predetermination is not a problem. The key aspect is therefore the character of the technology. When detached purposes in fact are relevant, the calculative approach may righteously be considered necessary and the generative approach a luxury. For non-trivial technologies, on the other hand, a lacking focus on generative aspects and a fixation of purposes to before-the-fact judgments imply a technology-enforced detachment from local aspects thereby inhibiting both the finding of local purposes and the achievement of effects related to these purposes.[47] The hierarchal view is insufficient because the relevance of effects depends, arguably, on the character of control. The dimensions are not independent, which is why they are not considered dimensions. They are two tensions.

The vertical tension (i.e., the character of controllability) must therefore be viewed over time. During design and system construction, a calculative approach is required, in order to know explicitly what to build. The design determines conditions for use and may thereby open up for possible situated usefulness, making the hierarchal view quite appropriate for the design case. Design is thereby not so much about handling a tension between opposing design approaches but more about setting a level of ambition for actually designing for situated controllability. However, a too strong focus on calculative aspects (i.e., an insufficient ambition to design for situated controllability) will for the case of use actively 'put the lid on' the possibility for situated usefulness (cf. the elevation of the delimiter between the calculable and contextual domains, the annexation of the activity domain by automation, illustrated in Figure 6.1). The fact that this lid also prevents human beings from discovering a need for adjusted or alternative purposes is what makes the calculative approach a self-fulfilling prophecy. It is much easier to appreciate what you get (calculative effectiveness and safety) than to comprehend what you miss (generative potentiality and resilience), until the missing aspects become tangible. When failing to realize that the situated aspect of usefulness is missing, the calculative usefulness becomes sufficient

[47]The implicit assumption of unambiguously valid values may actually be what Kahneman and Klein failed to disagree about when discussing what may be phrased as the issue whether to prefer rational or intuitive decisions over the other (Kahneman and Klein 2009). They never discuss according to what values the decisions are to be judged. For the statistical exercises in which heuristics and biases make people appear as poor decision makers (Kahneman's preference) there is an obvious value scale, defined by unambiguous mathematics, and in the fire extinguishing scenarios where expert people seem to outperform rational logics (Klein's preference) there is also an obvious desired outcome, a lack of casualties and an extinguished fire. But what about decision situations where the actions (supposedly dependent on more or less conscious decisions) actually alter the value scale according to which results are judged (cf. sensemaking as presented in ch. 9.4). For such decisions intuitive involvement becomes an end in itself, a categorical imperative, regardless of whether the result is objectively correct from a detached perspective (e.g., after-the-fact).

and the calculative approach validated. Hence, the main tension within the character of controllability is between the case of design and the case of use, a relation for which the meaning of the concept of usefulness is essential because communication between people is what connects the cases.

Furthermore, the categorization of purposes during design, into either achieving desired effects or avoiding undesired effects, might actually be somewhat irrelevant from an involved perspective during use. For generative controllability in the case of use, what matter are means for situated system control and for maintaining situation-, system-, and edge awareness, regardless whether controlled system properties are for achieving desired effects or for avoiding undesired effects. Generative controllability is about being in control of all possible effects, implying that means for control may coincide for both characters of utility. This is part of the rationale for stating that situated controllability must be an end in itself (e.g., ch. 14.1, addressing RQ2), as situated controllability appears significant for both generative safety (resilience) and generative effectiveness (potentiality).

If returning to the viscous surface metaphor for illustrating the character of usefulness, what would the optimal shape be for possibility of situated usefulness? The answer is probably: it depends! However, it is perhaps still possible to state a few general suggestions, at least a few suggestions about what seems unfavorable from the involved perspective of the human component. These following statements, phrased as metaphors using the viscous surface as a description of system characteristics, serves also as examples of the intended use of the model (R2.2). The intention is for the model to be used on a rather abstract level, as a framework for maintaining a rich meaning of situated usefulness, a meaning that keeps the involved perspective in view.

The case of normal uselessness coming from tight coupling and ambiguous purposes discussed above would be depicted as a technological system having a characteristics surface with peaks or a ridge aligned with the bottom edge, and insufficient or non-existing height on the upper half as a consequence. When real-life environments bring other conditions and purposes than predicted, an unfavorable mismatch between characteristics occur. The first [situated usefulness] conclusion to be drawn from these metaphors is then that the match between the shape of technology characteristics and the characteristics of the world in which the system is to be used probably is significant.

Another problematic characteristic for a technological system would be to have very distinct peaks, meaning a system highly optimized for specific aspects and lacking usefulness in between. Because such optimization arguably requires having fixed purposes according to design assumptions, these peaks would essentially collect against the bottom edge as well, implying normal uselessness, but the issue aimed to address with the sharp peak metaphor is its implications for practical use. An example of

technology with sharp and high peaks would be a system with explicit modes of fully automated operation for specific purposes. When real-life conditions and purposes align with a designed peak and the correct mode of operation is selected, a high level of usefulness is presumably experienced. The problem is that because of the complexity of the real world, natural peaks are rare, not easily distinguishable without subjectively filtering out certain features, it is a reality with richer characteristics compared to the stylistically exaggerated peaks represented by the operational modes. This kind of design implies ambiguities when choosing what mode to select and a lack of usefulness when mismatch occur. These peaks become thereby system performance edges possible to trip over. For an ability to avoid tripping the user must have good system awareness, presumably requiring to be very well aware of the subjective filters used to identify the peaks for which the operational modes have been designed.[48] If mode design filters differ from the users intuitive filters about activity structures, system awareness becomes a matter of deliberate and effortful application of high-level rational thinking, unlikely to be successful under stressful conditions and not contributing to 'gut-feeling' interpretations of interaction edges. The second conclusion is therefore that smooth (natural) characteristics probably are significant for intuitive system awareness that is suggested to be important for generative situated usefulness. Another way to put it is that because the controller must be a model of the controlled system (Conant and Ashby 1970) and because human beings in fact are intuitive creatures not possible to transform into fully rational machines, systems supposed to be controlled by human operators should match the analog and intuitively understandable characteristics that people can learn to understand.

The third conclusion is about making system characteristics match with environment characteristics, it is about design of technological systems. If the premise from chapter 5.2 can be accepted, that the physical world is more appropriately described as a deterministic (hard) system than the social world, then it is possible to state that the more physically bounded a technology is the more appropriate is a calculable approach to controllability. When designing a trivial technology, for example an angle iron, it is perfectly fine to maintain a calculative approach, once and for all define its purpose or purposes and derive design specifications by use of hard-systems thinking and mathematically describable laws of physics. When used, the properties of the angle iron do structuralize the work, of course, but not actively, and it may of course be used for other purposes than those governing the design, possibly resulting in a change of its considered usefulness. However, because the angle iron is a passive artifact, all activities in which it participates and the resulting effects are naturally

[48]This statement is describing what also could be phrased as the necessity for a match between designer and user mental models (e.g., Norman 2002).

attributed to its applicator (i.e., its user). For technologies more complex, it is not that simple. Systems with machine workings take on a more active role and may thus on their own create effects, although for mechanical non-computerized machines their effects are still considered fundamentally governed by physical aspects, implying that human beings do the thinking and provide physical control input to the systems. Mechanical technologies are by their nature restrained to the physical domain thus only capable of structuralizing physical aspects of our world and only indirectly by physical structure shape our interpretation of it.

If the other premise can be accepted as well, addressed in chapters 6 until 9, but discussed particularly in chapters 5.4 and 6.3, the premise that computerization make technology become involved in social aspects to a much greater extent than traditional mechanical systems. If that premise can be accepted, then the natural demarcation between thinking human beings and physically performing technology becomes blurred. Because of their logical information manipulating capability, computers can think by themselves in a manner that resembles our rational kind of thinking (i.e., system 2). Consequently, computerized systems can take on an active role also in our mental activities, thereby structuralizing the way we think, much more directly than any kind of technology have ever done before. The third conclusion is therefore that computerization makes a deliberate, conscious, and explicit matching of system properties to environmental conditions and user properties (i.e., human qualities), an increasingly urgent issue.

Computerization, as in automations, templates, filters and modes of operation, systems that actively take part in situated activities by application of their predefined behaviors, become simultaneously the strongest ever enforcer of calculative judgments and obscurer of a potential lack of situated aspects. Computers, the kind of technology that makes it necessary to explicitly adopt a generative approach when designing them, is also a kind of technology that frames the design approach to calculative aspects because it is a kind of technology that shapes our thinking. When the obscuring of situated aspects become strong enough to make indications of their necessity considered a problem, which for example happens when usefulness is judged against similar models as those obscuring the alternative aspects, a vicious circle of reduced concept richness is entered. The present quest for situated usefulness and edge awareness aims to be a remedy for this unfavorable situation that is characteristic for the contemporary view of usefulness. In order to use computerized technology to its full potential for the benefit of humankind, a rich meaning of usefulness is required, which is a meaning that includes the involved perspective required for true human values.

10.3 The case of design and concept meanings

The character of control for a technological system in the case of use is the result of the approach to (i.e., the attitude towards) situated controllability (the first stage in R1.2.2, the character of control) during design. This fact implies that the meaning of essential concepts is crucial, simply because these meanings govern design intentions. Concept meanings are, obviously, crucial for all social activities. In the present context, however, the main concern is the interpreted meaning of concepts in the case of design. Design is a creative process and a social activity requiring people to communicate ideas, views, and values Future designs (the design result, the second stage in R1.2.2 – the character of control) depend on interpretations and evaluations of present designs, which implies communicating views and values. If the interpreted meanings of concepts used in such communication lack the involved perspective, the focusing on the detached perspective becomes a self-reinforcing trend. With concept meanings deprived of situated aspects, it will no longer emerge a need to talk about the involved perspective, and the reduced richness of concepts becomes established without being challenged. Judgment and reason about usefulness has then become instrumental (Weizenbaum 1976). However, because the lost involved perspective is the natural human perspective, it is of utmost importance that all concepts describing human qualities and virtues make the involved perspective comprehensible. The trend appears, however, be more like the above. Following our tendency to consider computerized technology as thinking objects, it seems that we want to use concepts traditionally reserved for human aspects also for technological properties. This development is apparently going on for several human-related concepts such as intelligence, awareness, consciousness, and autonomy. Because these concepts are also used to describe the involved perspective and ourselves, the approach to use them for technological properties makes us consider ourselves from a detached perspective. The result is that the distinction between qualitatively different aspects becomes blurred.

Chapter 15 is an elaborate discussion about implications of inappropriate use of essential concepts and about shifting concept meanings, which largely is a recapitulation of the article “Autonomous technology – sources of confusion: a model for explanation and prediction of conceptual shifts” (Stensson and Jansson 2014a). The article discusses the development with reduced concept richness explicitly for the concept of autonomy, thereby describing what is believed to be an important insight made while theorizing about usefulness. The discussion is based on the four identified aspects that make up the third result (R3) of phase one.

The first aspect (R3.1) is the tendency of reduced concept richness introduced above. This reduction was identified as plausibly driven by two natural human desires, the desire for comprehensibility and the desire for

predictability. Explicitness and concreteness, commonly accomplished by adopting a detached perspective, tend to make things rationally comprehensible, clearly a desirable quality. Predictability makes also the otherwise incomprehensible future appear comprehensible, in the sense that predictability makes it known. The detached perspective facilitates also the use of objective measures and rigid mathematics, which is thoroughly reassuring. However, it seems there is a tendency of having an exaggerated faith in predictability and objectivity, which is the second aspect (R3.2). The third aspect (R3.3), a combination of the first two, is showing as an exaggerated strive to accomplish predictability, presumably because the desire for comprehensibility and the exaggerated faith in objective predictability make the desire for predictability appear possible to satisfy. It is, however, possible to interpret the desire for and exaggerated faith in objective predictability as a desire to escape from responsibility (R3.4). With perfect objectivity and predictability, there is no need to take responsibility, simply because there is an unambiguous truth. However, such an escape implies giving up on human autonomy because to accept the view provided by supposedly objective models is to be heteronomous.

All in all, these four aspects make up what here is called the vicious circle culture (R3), a culture identified as a plausible explanation for the persistence of technical rationality and the contemporary view of usefulness that mainly focus on calculative effectiveness and safety. As such, the contemporary view implies stereotypical or illusory effectiveness and brittle or false safety, lacking situated potentiality and resilience because system workings are forced to comply with simplified and limited models. The vicious aspect of the culture is that the reduced concept richness makes us unable to see or communicate that there is a lack of richness and our desires for comprehensibility and predictability tend to make us interpret the lack of insight as the ideal. By reducing the richness of essential concepts, we are putting our heads inside a box that we are unable to think outside of.

– Phase two –

11 Case studies, exploring RQ2 and RQ3

In phase two, an empirical exploration of situated usefulness was conducted, based on external data from documented incidents. In relation to phase one and to the context of discovery, this exploration served also the purpose of providing additional input to the theorizing process (ch. 2.2.2 & 2.3). The theories were generated through retroduction (or abduction) to plausible explanations (i.e., theoretical hypotheses accompanied with reasons why they are plausible) in a parallel process of deduction from external theories and induction from personal experiences, complemented with enhanced understanding from scrutinizing the following cases. However, phase two serves also the purpose of empirical corroboration, thereby constituting an effort within the context of justification as well. The second and third research questions were formulated for all of these purposes, one explicitly for the case of use and one for the case of design, they were:

(RQ2) **The case of use**:
What is the role of a human user when using potentially useful technology and how does this role affect actual usefulness?

(RQ3) **The case of design**:
In the case of design, how does the view of the human role in the case of use affect the usefulness of the designed technology?

11.1 Case study methodology

Four major and three complementary cases were studied using a qualitative methodology with the dual purpose of enhancing the understanding of situated usefulness and corroborating the theoretical contribution of phase one. The overall structure and principles for the particular kind of method used for qualitative data analysis was presented in chapter 2.3.2. Presently, the purpose is to briefly discuss why these particular cases were selected, introduce them and their sources, and explain the specific case study structure used to address both RQ2 and RQ3 with the selected cases.

In order to explore RQ2 – The case of use, two different kinds of technological systems were studied, and in order to explore RQ3 – The case of design, two timeframes were selected. The main cases were two airliner

incidents, one incident from the past (SAS SK751 in Gottröra, 1991) that actually ended quite happily, and one present case (Air France AF447 in the Atlantic, 2009) that was fatal, as well as, two nuclear power plant incidents, one from the past (Forsmark, 2006) that as a matter of fact also went by without any significant consequences, and one present (Fukushima, 2011) that became a disaster (still going on). As a complement, three future technology scenarios have been scrutinized, treated in principle as one fifth case, thereby functioning as one additional kind of technology as well as one additional timeframe. The layout of these five cases in relation to the exploratory research questions are shown in Figure 11.1.

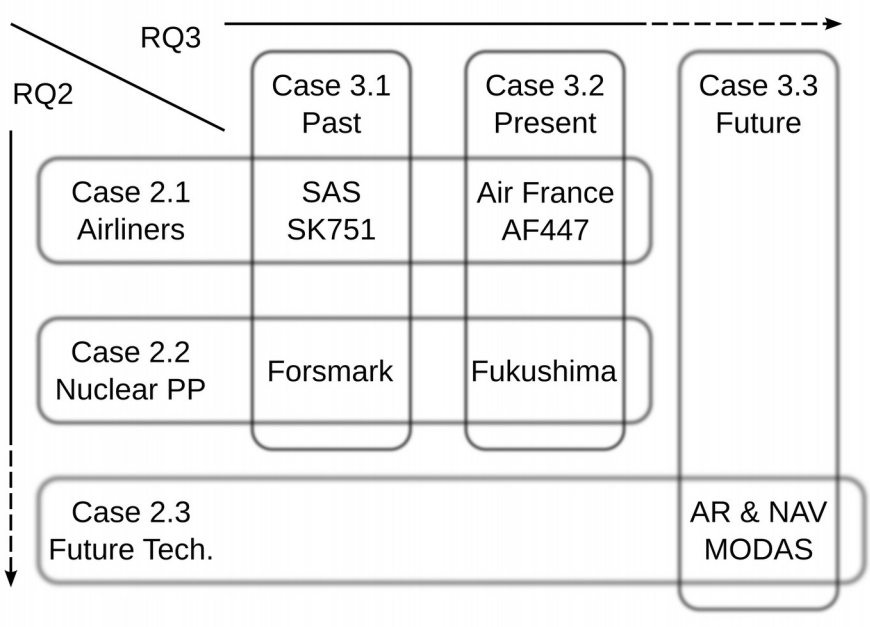


Figure 11.1: Relative layout of cases studied

The four primary cases are all studied by reference to incident investigation reports, produced by independent and formally appointed authorities. The SAS crash in Gottröra was investigated by the Swedish Board of Accident Investigations, the Air France crash was scrutinized by the French Bureau d'Enquètes et d'Analyses pour la sécurité de l'aviation civile, and Fukushima was investigated by The National Diet of Japan Fukushima Nuclear Accident Independent Investigation Commission (referenced in the corresponding chapters). These organizations are by virtue of their official status considered quite objective and trustworthy sources, and their reports the most reliable information available about the studied cases. The exception to this rule of assumed objective trustworthiness would be the report about the Forsmark incident that was not considered severe enough to

merit an independent accident investigation. The incident was instead investigated by the analysis group at the Swedish Nuclear Training and Security Center (KSU),[49] self-proclaimed as being independent, but still part of an organization jointly owned by the Swedish nuclear power plant operators. Consequently, the information available from their rather brief reports is considered, possibly, slightly biased, which also is discussed further in the presentation and analysis of the case.

The choice of two airliner incidents as one kind of systems to scrutinize was partly because of domain familiarity and access to detailed analyses, but also because they are cases of general interest as well as cases of high representativeness. Airliners are aircraft, obviously, thus suggesting some kind of applicability and ensuring some sort of relevance for conclusions drawn on behalf of general aeronautical experiences. Predominantly, airliner incidents are also thoroughly investigated, resulting in detailed reports that in addition seldom become restricted or classified, as might happen with military aircraft incidents. Furthermore, airliner cases are representative in several ways for the present research context. Because most people can relate to airline traveling and fear airliner crashes, they become normal failure cases of high interest. Yet they are examples of advanced real-time dynamic systems with immediate consequences. In short, airliners exhibit many of the illustrative qualities of the extreme case argued as a significant reason for studying military fighters, while simultaneously airliners are by most people not considered particularly extreme. The connection between issues in extreme situations and similar ones in these cases should therefore be rather straightforward.

The choice of nuclear power plant systems was also due to multiple reasons, some reasons equal to those considered relevant for selecting airliners. Consequences of nuclear power plant accidents are of general interest because they affect many people, thereby ensuring some level of relevance to the public, and there tend to be detailed incident reports available from independent authorities. Furthermore, as nuclear power plants are highly automatized because of their nature, just like unstable fighter aircraft, there is, supposedly, some kind of applicability for insights about control issues from the aeronautical domain.

In fact, the high level of automation is a common denominator for all the main cases, and there is a connection to safety in all cases. Both airline traffic and nuclear power production are considered ultra-safe systems, a safety that largely is attributed to an extensive use of advanced automation and other model-based behaviors, technologies presumed to assist involved people in maintaining safety (at least keeping people from making dangerous errors). Yet are all these cases examples of system breakdowns. Furthermore, they are all examples where an exaggerated belief in model-based behavior

[49]In Swedish, Kärnkaftsäkerhet och utbildning

actually may be concluded as the cause of the accident, despite a significant lack of such conclusions in the official view presented by the investigation reports (except for the Fukushima case, attributing the cause to the organizational model). Striking indications of a calculative culture, is it not?

As a bonus, the two past cases have in addition a touch of local relevance by virtue of geographical proximity. The crash site in Gottröra, where SK751 came down, is located merely 30km from Uppsala city center (and from the university at which the present research has been conducted), and the nuclear power plant at Forsmark is only 68km away. This fact was, presumably, most significant for selecting the Forsmark case which, however, happened to facilitate a late update because, actually while analyzing the event and writing the chapter, a similar incident took place. There is more about this event in the analysis of the Forsmark incident (ch. 12.2).

The methodological idea is that aspects identified as relevant for explaining usefulness can be generalized from these specific cases because the case study structure allows for comparison between different types of technology as well as between different timeframes. As such, the structure becomes, on a superficial level, a cross-case display with simultaneously case-ordered and time-ordered cases (Miles and Huberman 1994, pp. 187, 200). While both types of cases in a sense regard real-time kind of systems (vehicles and process industry), they are still rather different in character. Aspects of usefulness found both in airliner systems and nuclear power plant systems (as well as in future systems) should thereby be possible to consider quite general. Analogously, aspects of usefulness found in both past and present cases (as well as in future scenarios) should be possible to consider more than just random occurrences. In particular, the different timeframes can be used as a reference for the case of design because designs are produced in the past and used in the present, or rather, produced in the present and used in the future. This comparison between timeframes means that if aspects relevant for describing issues with usefulness in past cases still are relevant in present (and future) cases, they are likely to belong to an established culture (an underlying social mechanism) either not considering these undesired effects or a culture unaware of their relevance.

The cases were scrutinized, initially with focus on RQ2 – The case of use, although by use of the theoretical contribution. This was done essentially by keeping the following two questions in mind while reading the reports:

- Did the crew/team see the performance edge coming?
- Did they have situated controllability, that is, did they have means to intervene?

12 Cases from the past and the present

12.1 SAS, SK751, December 27th, 1991

Scandinavian Airlines System (SAS), today called Scandinavian Airlines or simply Scandinavian, is an international airline company formed in 1951 as a conglomerate consisting of the three Scandinavian flag-airlines still partly owned by the Swedish, Norwegian, and Danish governments. By the time of this accident the company (with their subsidiary companies) had for a while been in a transition phase, from the situation of being a governmental institution practically having a monopoly for public air transport, to the situation of a more competitive free market. The old role as an approved authority may have had some impact on the company culture in a way that contributed to the accident, but there were also a number of technological issues played their parts. What happened was the following.[50]

Friday morning at 08:51 local time on December 27th, 1991, SAS flight SK 751 from Arlanda just north of Stockholm, bound for Copenhagen, lost both its engines and crashed outside the small village of Gottröra close to the border of Uppsala municipality, merely four minutes and five seconds after takeoff. Miraculously, not a single one of the 129 people on board was killed in the crash, although some were injured for life and others still suffer from post-dramatic stress. Nevertheless, because this was just after Christmas, the accident, which easily could have become a tragic disaster, is sometimes referred to as The Christmas Miracle in Gottröra.

The weather this morning was, as it might be in these parts at that time of the year, just below freezing (-0°C) with a dew point at -1°C and light snowfall,[51] but still with good visibility. There were a few clouds at 600ft (~180m) and a slightly more covering layer at 800ft (~250m). The aircraft had arrived from Zürich late the evening before, a flight that took place at high altitudes with temperatures between -53°C and -62°C, and afterwards it was parked outside during the night with about 2500kg leftover fuel in each wing tank. Wing tanks are usually, as for this aircraft, what is called integral tanks, designed with the sealed wing surfaces as its container walls, thus shaped as the wing itself. The contained fuel collects thereby naturally in the

[50]The accident investigation report (SHK, Swedish Board of Accident Investigations 1993), is in this chapter (ch. 12.1) referred to by SHK only.

[51]The dew-point is always close to the current temperature when there is precipitation, but note that this particular combination constitutes nearly optimal icing conditions.

lowest part of the tanks, which happens to be near the wing roots located in front of and precisely in line with the engines. As it was, this aircraft happened to have a significant amount of very cold fuel in the wings (way below freezing) and the weather conditions were practically optimal for ice to develop, so obviously it did.

Before takeoff the aircraft was de-iced with extra care, even the undersides of the wings were ordered treatment after a dialog between the captain and a technician. After de-icing the crew taxied to runway 08 (heading east, ~080°) and according to standard procedures they performed what is called an immediate (rolling) takeoff without stopping at the beginning of the runway. Takeoff proceeded normally until rotation, which is where the aircraft has gained enough speed to lift its nose and actually take off, when a slight rumble was noticed, followed by an unidentified sound also registered by the Cockpit Voice Recorder (CVR) as a weak humming. What happened was that, evidently, the de-icing had not been satisfactory performed after all, and when the wings as always were slightly bent when the aerodynamic forces start to work and the aircraft begin to fly, the stiff clear ice cracked and chunks of it flew off. Some of these loose pieces of ice were sucked into the engines located straight behind the wing roots.

About 24 seconds after rotation, at an altitude of 1124ft (~343m) within clouds, the right engine surged, the compressor stage stalled[52] and the engine began pumping. This is a turbine engine malfunction typified by loud bangs and possibly with flames “pumping” out of the engine air inlet and jet stream exhaust. When the airflow through the air inlet and compressor stage for some reason becomes disturbed, the fan blades begin to stall and the compressor can no longer uphold sufficient pressure to match the higher pressure within the engine, which thereby will drop quickly. With lower pressure inside, the compressor will usually recover directly, but if the disturbed conditions prevail, it will obviously stall again. The result is that the highly pressurized extremely hot and burning air and fuel mixture from the combustion stage “pumps” back and forth through the engine causing the loud popping sound and the burning air to appear. Disturbed air may come naturally into the engine air inlet, for example if the aircraft passes through the jet stream of another aircraft in too close proximity. For such a case, the engine will probably recover as quickly as there is undisturbed air again.

However, if the compressor is malfunctioning because there has been some kind of Foreign Object Damage (FOD), undisturbed air may not be sufficient. The most common cause of FOD is perhaps from birds that are 'swallowed' by an engine, or from things being splattered up by the wheels

[52]Stall for compressor fan blades is principally the same as stall for aircraft wings, the aerodynamics of the wing or the fan blade does not work properly because of disturbed or otherwise inappropriate airflow, this results in loss of lift for a wing and in loss of pressure behind a compressor. The Air France disaster case (ch. 12.3) includes stalling wings.

when rolling on the taxiway, or from ice.[53] If the damage is not too severe, the engine may still be able to work fairly well, but with a reduced maximum thrust. What therefore has to be done is quickly to stop the pumping and find the maximum thrust available, and never exceed that limit. Pumping implies extreme strain on the engine and it will not hold together for long. Sooner than later, it will begin falling apart and loose parts will damage the engine further. The right engine broke down completely after 51 seconds of compressor stall and pumping, the left engine 2 seconds later.

That is, the left engine did unfortunately also surge, the compressor began to stall 30 seconds after the right engine and 64 seconds after rotation, but this is something the pilots never registered, and the left engine did not last for more than 14 seconds. At that time, the aircraft had reached an altitude of about 3200ft (~980m) and the pilots had initiated a left turn. Parts from both engines have been found on the ground just below this position, indicating the devastating strength of the pumping. With both engines irrevocably out, the electrical system also started to fail and the digital flight instrument screens on the left side in front of the captain went out. He never tried to cross couple the information from the right side, something that would have been possible although it probably made no significant difference for the accident outcome. The consequence was, however, that the captain had to rely on the smaller reserve instruments for the emergency landing.

The crew noticed the first engine malfunction all right, at least the co-pilot identified it correctly because he said "... believe it is compressor stall" (in Swedish).[54] They deduced from the instruments that it was the right engine, although the captain said later that he had problems reading the digital displays due to vibrations and the quickly changing values. The captain reduced the right throttle slightly, although not enough to stop the engine from pumping. The co-pilot said afterwards that it was not until both engines were out he noticed the warnings on the engine instrument panel and that the exhaust temperatures were much too high. Thirteen seconds later the fire alarm went on for the left engine and he triggered the fire extinguishing system. The fire alarm went off after another 26 seconds.

There was also some involvement from informed passengers. The flight attendant sitting at the aft folding seat was for instance alerted by a pilot captain on a private trip about the right engine having compressor stall and she tried without success to reach the cockpit with this message through the intercom. Then there was in addition this uniformed pilot captain sitting on the second row who realized that the crew had severe problems and went to

[53] Aircraft engines are designed to withstand water amounts below a certain limit and birds and other objects below a certain weight. These particular engines were designed to cope with "soft objects" up to 2.14kg (SHK, p. 24), but there is always the risk of having the compressor fan blades damaged in such a way that they either brake and scatter or otherwise damage the engine linings.

[54] See footnote 55, p. 178, about CVR transcript quotations

the cockpit to offer his assistance (the door was open), something that clearly was (and is) against the rules, but in this case it appears to have been crucially beneficial. There is no doubt that the successful emergency landing, the fact that the final crash into trees and on to the field was as smooth as it was, largely was pure luck. Anything can happen when 55000kg hit the ground at 220km/h (121 knots). There can be no guarantees when spot landing a damaged airliner in wooded terrain, but this assisting captain seems to have been a great asset. He was to begin with welcomed by the crew, the co-pilot handed him the emergency checklist, and the flying captain tasked him to try to get the Auxiliary Power Unit (APU) started in order to fix the malfunctioning electrical system. The assisting captain was from then on heard coaching the crew all the way down to impact, mostly by keeping them focused on actually flying the plane and helping them spot the field when they came below clouds at about 900ft (~280m). On the CVR he is heard repeatedly saying, “Look straight ahead”, “Keep looking out” (in Swedish). He also extended the flaps gradually (which increases the lift at low speed and help reduce the speed further), so when the captain said something like “(flaps, eh eh)” he answered with emphasis “Yes we have flaps, we have flap[s], look straight ahead, look straight ahead!” (in Swedish, the bracketed plural 's' is ambiguous throughout the report). Finally, about seven seconds before impact, the assisting captain called out “Choose a spot, right, right, right, right, right,… right, right,... turn right, turn right” and shortly later “Yes straight ahead there, straight ahead there, straight ahead (straight towards the woods)” (in Swedish).[55]

What happened then is, as it goes, history. Most of the right wing was torn off from hitting the trees before reaching the open field that they were aiming for, and the aircraft started to lean right. The last registration before impact (-1s) says 198km/h (107 knots) and 19.7° right bank. The aircraft hit the conveniently downwards sloping and completely frozen ground on the open field with its tail section first and a 40.1° right bank, and it skidded for 110 meters on the few centimeters of snow that was covering the field, before at last coming to rest. The fuselage was broken into three pieces and 6900 liters of jet fuel together with 100 liters of hydraulic fluid was spilled, but there was no fire (cf. Figure 1.2, p. 13). All but four people out of the 129 on board managed, by themselves, to get out of the wreckage. Eight were severely injured, 84 got minor injuries, and 37 were registered as having no injuries at all. Things could definitely have been much worse!

[55]The “quotations” (i.e., my translations) are from the transcript of the CVR presented in the report, where sections within parenthesis are stated as highly uncertain. The co-pilot and the assisting captain are Swedish and they spoke mostly Swedish during the incident, the captain Pilot-in-Command is Danish but apparently rather fluent in Swedish (he was talking “Swedishly” earlier in the transcript), indicating that the mixed Swedish-English conversation was completely normal thus assumed not to have been a problem. Such mixed language was at that time (and is probably still) quite normal in the Swedish aviation community, also on official flight-control frequencies, a fact that I can confirm by personal experience.

12.1.1 Accident investigation findings

The technological system of concern is an airliner of type Douglas DC-9-1 (McDonnell-Douglas MD-81), which is a descendant of the DC-9 airliner originally designed as early as in the 1960s. This particular aircraft was however quite new. It was built earlier the same year it crashed, delivered to SAS on April 10th, 1991, but as a short-haul airliner for the three countries of Scandinavia that makes up a relatively small geographical region, this trip was already to be its 1273rd flight[56] (it collected a total of 1608 flight hours). The aircraft type is characterized by a long narrow body (45.1m long, five seats abreast), slender sweptback wings (32.9m wingspan), two engines far back at the sides of the fuselage, and a T-tail (a high mounted horizontal stabilizer) due to the engine placement. The relatively recent production date made this actual aircraft to be fitted with a number of computerized subsystems of which at least one had a vital role in the accident. The properties of these digital sub-systems are therefore of particular interest.

For normal operation with a fully working system, the main flight control information is displayed on the Electronic Flight Instrument System (EFIS) of which there are two identical, one on each side of the aircraft. Its major interface consists in two color displays, one for each of the two pilots, on which synthetic images of the central flight and navigation instruments are presented. Cross coupling of the displays is possible in case one of the systems would fail. To restart the EFIS system takes one to six seconds.

The Digital Flight Guidance System (DFGS) is a doubled (redundant) autopilot and navigation system, intended to reduce pilot workload and it is certified for use after reaching 200ft (~60m) during take-off. Within the Digital Flight Guidance Computer (DFGC) several sub-systems with different functions are integrated and system feedback is displayed on the Flight Mode Annunciator (FMA), a small display of which there are two, one positioned in front of each pilot. For this incident, the engine control automations are of primary interest.

The engines are ultimately controlled by one Fuel Control Unit (FCU) each, which, depending on input from throttle, engine parameters such as the different rotor-shafts rpm measures, combustion chamber pressure, and the outside air temperature, dispense fuel to the fuel injectors. From the pilot end, there are the “traditional” throttle levers (one for each engine) and a common Engine Pressure Ratio (EPR) panel placed in the central cockpit console, thus reachable from both pilot seats just like the throttle levers. The EPR panel gives input to the Thrust Rating Computer (TRC) integrated in the DFGS and it has a number of major working modes such as 'max takeoff', 'normal takeoff', ′go-around', and 'climb'. The TRC is in turn providing input to, among other things, the Auto Throttle System (ATS)

[56]This number is a bit ambiguous. The report states 1272 cycles but does not mention whether or not this final trip was considered a complete cycle. Thus it may have been its 1272nd flight.

essentially working on both throttles consistently by actually moving the throttle levers with servomotors. The pilots can always overmaster the ATS, by simply shifting the levers with force, or by turning off the ATS with switches on the levers. There is also a certain amount of engine-interdependent ATS functionality, for example when the engines, within a certain interval, are synchronized to have equal thrust settings. During takeoff the ATS becomes automatically switched off when the measured air speed reach 60 knots (~110km/h), and a setting called CLAMP becomes enabled, indicated with CLMP on the FMA and a fixating of the throttles in their present position. This fixation is released as soon as a new thrust mode is selected.

There is also an Automatic Thrust Restoration (ATR) system and an Automatic Reserve Thrust System (ARTS) integrated into the DFGS that work independently of each other. The ATR is designed to automatically disable CLAMP and enable ATS so it can move the throttle levers forward (increase the thrust) if, under certain conditions, engine problems are identified. The rationale for this automation is that many airlines use what is called noise abatement thrust cutback procedures, where thrust is reduced during climb directly after takeoff to reduce the noise footprint. Because this procedure also makes the aircraft more sensitive to a sudden loss of thrust, the ATR is explicitly designed to revoke quickly and automatically the thrust reduction in case of problems. SAS was not using these procedures at the time and according to the accident investigation report, the company was actually unaware of the fact that ATR was installed in the aircraft. Furthermore, the ATR works covertly and cannot be controlled directly, it is automatically armed when the aircraft is in takeoff mode above 350ft (~105m) and when the actual thrust on both engines is below 'go-around', it is automatically activated when actual thrust or rotor shaft rpm differs too much between the engines. Activation means triggering the ATS in a situation where the thrust levers usually are locked in CLAMP mode. There is no explicit indication of ATR being activated. The pilots must deduce this from the fact that the CLMP indication on the FMA changes into EPR G/A (go-around), that the T.O. (takeoff) button-lamp goes off on the EPR panel, and that the GA (go-around) button-lamp on the EPR panel goes on, and, from the fact that the throttle levers begin to move forward as a result of the ATS being automatically engaged. The ARTS system has a similar purpose as the ATR, to provide extra thrust if engine problems are discovered, but ARTS does this differently, by directly increasing the fuel flow to the FCUs by opening electrically controlled valves, not affecting the throttle levers. There is, however, feedback on the instrument panel and the system can be directly controlled by a switch.

The icing problem is especially problematic for this aircraft design, which is well known by system operators (airline companies and control authorities). Similar incidents had been reported several times before. In

particular, the compressor stages of the engines on a DC-9-51 belonging to Finnair from the neighboring country of Finland was severely damaged this way, and their report from 1985 stated the clear ice problem as “the most difficult systematic threat to flight safety today” (SHK, p.60). The high impact of the problem on global flight safety at that time in history was because there were many of these airliners in operation, almost 2000 aircraft of the different DC-9 versions have been built. Their high sensitivity to icing is due to the combination of fuel tank placement and engine mounting. The DC-9-1 version has one central body tank that extends about 1.6 meters into the wings feeding both engines, and two wing tanks, each one normally feeding the engine on the corresponding side. The central tank is usually emptied first, followed by the wing tanks, which implies that excess fuel, often very cold from having been at high altitudes, resides mainly in the wing tanks, and their shape make the fuel collect towards the lowest and innermost parts against the bulkheads between the wing tanks and the body tank. With a significant amount of very cold liquid within, the wings will continue to be cold and ice can develop rather quickly during a stop even at warm places, where ground personnel in addition can be assumed to have less experience from problematic icing conditions. When there is a cold surface and moisture in the air, ice develop, and under certain conditions the ice may become completely clear thus very difficult to identify. To compensate for this difficulty, the aircraft was fitted with markings that become visually distorted and cords that stop moving when stuck in ice, in the critical regions to help spot the presence of clear ice. Furthermore, only two months before this accident, on October 26th, SAS issued a FLIGHT DECK BULLETIN/WINTERIZATION stating, “It is the P-i-C [Pilot-in-Command] responsibility to check the aircraft for any ice and snow that may affect the performance”. Moreover, in a special section labeled CLEAR ICE, the bulletin stated, “Although the awareness within Line Maintenance is mostly good, the responsibility again leans on P-i-C that the aircraft is physically checked by means of a hands-on check on the upper side of the wing. A visual check from a ladder or when standing on the ground is *not* enough” (SHK, p.62, both quotations, brackets added, emphasis in original).

The engines, that have air inlets of about 1.2 meters in diameter, are fitted behind and slightly above the wings, mounted on pylons from the sides of the aircraft body making the turbine centerlines be located about one meter out from the fuselage. This makes in turn the outer border of each engine inlet horizontally align perfectly with the inner border of each wing tank. The slight deflection thus required for ice braking loose from the innermost area of the wing tanks to end up in the engines is minor. Furthermore, during takeoff directly after rotation the angle of attack[57] is at its highest because the

[57]Angle of attack is the angle between the aircraft's length axis (or the wing's baseline) and the meeting airflow. The larger the angle of attack the higher the lift and drag, until the angle is too high making the wing stall, deprive the wing from producing lift, leaving only drag.

aircraft is then flying at its lowest possible speed, making this the situation when the airflow from the wings pass the closest to the engines. Turbine engines at high thrusts work also like gigantic vacuum cleaners, swallowing extreme amounts of air per second, and anything else that comes in too close proximity.

This eventful December Friday was also on the verge for SAS to become a double disaster day. Another MD-81 that also had spent the night outside, SK 483 to Oslo, was de-iced by the same personnel but inspected by another mechanic, took off 18 minutes after SK 751[58]. After landing in Oslo one passenger informed the captain that he had heard unusual sounds during takeoff and saw clear ice on the wings. Inspection found that the left wing was covered to 20% by clear ice and the right one to 30% beginning about 1.5m from the fuselage. After a brief inspection of the engine air inlets the plane was flown back to Stockholm Arlanda, where a more thorough inspection found that five fan blades on the left engine had soft dents on their front edges and that they had to be replaced. Evidently, this was a very close encounter.

The report concludes that the accident was caused by insufficient instructions and procedures within SAS for ensuring that clear ice is removed from the wings before takeoff, in spite this being a well-known hazard. Thereby the aircraft came to take off with clear ice on the wings, ice braking loose during rotation, flying off and being sucked into the engines. The ice chunks damaged the compressor stages of the engines, which resulted in compressor stall and pumping, severe enough to destroy both engines. The report states as contributing factors, insufficient training of pilots to identify and handle compressor stall and pumping, and, the Automatic Thrust Restoration (ATR) system, which, without pilot awareness, increased the throttle settings and made the pumping worse.

12.1.2 Interpretations according to present frameworks

While the main cause of the accident was organizational according to the investigation report, focus for the present analysis is on the fact that there were automated technologies that made the hazardous situation significantly worse. First, a little closer look into the icing problem. This particular aircraft design (highly sensitive to icing) is obviously rather impractical for northern climates with periodically reoccurring extended periods of icing conditions (i.e., long cold winters), but still usable, if sufficient procedures are properly followed. There were in fact pieces of information in manuals about the icing problem, and there were some regulations in effect, all while the report highlights quite a substantial number of shortcomings and

[58]As a slightly obnoxious reflection it is possible to ask oneself why SK 483 actually was allowed to take off, as this was 14 minutes after SK 751 had crashed, an event that ought to have triggered some extra caution and analysis, should it not?

suggests several improvements. There were procedures that *should* have been known and implemented by SAS, procedures that in practice ware done so only to a limited extent. SAS was simply, from a holistic organizational perspective, not fully updated and sufficiently concerned about the icing problem to handle it satisfactorily. For this non-complacency to apparently desirable international standards, the inherited company culture as a former governmental air traffic authority could perhaps be blamed. In fact, the report points out with emphasis that the idea of having a self-controlling responsibility for companies presupposes that regulating authorities make sure that the companies actually have functional and working self-controlling procedures in place (SHK, p.77).

Ultimately, it is the captain in command that has the responsibility to make sure the aircraft is ready for takeoff, this was made quite clear, not least by the aforementioned winterization flight deck bulletin, but in practice, certain checks have to be performed by ground personnel. There were in fact regulations in effect (although reported as inconspicuously conveyed) stating the responsibility of ground mechanics to check for clear ice by touching the critical area on both wings with bare hands. Although, for SAS ground personnel, this was a task for which they lacked appropriate equipment, such as ladders or the like, so there was ice (SHK, p.77).

The present analysis will however not settle with the argument that this accident would not have happened if these regulations and procedures actually had been followed (that is considered a too calculative approach), but instead consider the situation as an emergency that might have occurred anyway. Jet engines can in fact get ice or other foreign objects into their inlets, and have their compressor fan blades damaged, regardless the amount of precautions. Certain unfavorable environmental conditions may befall, or there may be other causes for a (sub-) system breakdown. Let them be low-probability events, yet for them to occur is not impossible. Whatever reason for why such a hazardous situation might occur, it will be of crucial importance to have the ability to control the malfunctioning technological system (i.e., the engines) in a non-routinely but still insightful manner. What if there had been birds damaging the compressors instead of ice? Like for the astonishingly similar accident, accordingly nick-named 'The Miracle on the Hudson' (National Transportation Safety Board 2010), where US Airways Flight 1549 on January 15th, 2009, took off from La Guardia Airport in New York city, got some birds (allegedly several large birds) in the engines and made an emergency landing on the Hudson River completely without casualties. Can the presence of birds be blamed on flaws in organizational procedures? Probably not! Therefore, a more constructive (and generative) approach must be to consider the emergency with malfunctioning engines merely initiated by the presence of ice (a presence that, by all means, might have been caused by organizational flaws), but consider the crash caused by

something else. Plausibly, the plane crashed because both engines were destroyed, which perhaps could have been avoided.

> Nothing indicates that the engines had any other damages when the pumping started besides the limited damages induced [by ice from the wings] when the aircraft took off. These damages were not more severe than that the pumping in the right engine probably would have stopped if the thrust setting had been reduced sufficiently much. Pumping would probably not have occurred in the left engine at all, if the initial thrust setting had been kept during climb. With a sufficiently reduced thrust setting for the right engine and a maintained thrust setting for the left engine, the engines would probably not have been destroyed. The aircraft should then have been able to return [to Arlanda] for landing (SHK, pp.74-75, my translation,[59] brackets added).

If that is true, that the engines could have been saved by an appropriate handling of the thrust settings (the throttles) and thereby allowed the aircraft to return safely for landing at the airport. Obviously, this must mean, contrary to what the report itself states, that the accident in fact was not caused by the presence of ice, but by whatever made the engines brake down after they were damaged. The interesting question therefore becomes, why were the pilots unable to keep the engines running? The answer is arguably (in part) also there in the report, but listed as contributing factors.

The report lists two contributing factors to the accident (SHK, p. 86). The first factor was that the pilots were insufficiently trained to identify and remedy compressor stall and pumping. The second factor was the ATR system, a system that became activated and that increased the throttles without the pilots being aware of neither its existence nor its function. The ATR system was at the time unknown to SAS. Both these issues regard control of aircraft subsystems consisting of the two jet engines.

So, why did the pilots not stop the pumping that destroyed the engines? One plausible explanation is that they were not sufficiently aware neither of the fact that the engines actually were pumping, nor of the severity of the situation. By considering R1.1.1 (Figure 5.1, p. 99) and R1.2.1 (ch. 7.2 – Character of uncertainty) and the design of this particular aircraft with its engines located about 40 meters behind the cockpit, it appears reasonable to state that the natural physical feedback of engine workings may be minor. The pilots are then for evaluation of the situation confined to actively consult instruments only capable of conveying feedback they are explicitly designed to convey, which usually concerns only information relevant for normal operation. Without hearing the bangs, or feeling the tremors in the fuselage, or seeing flames pumping out of the engine inlets, which are all natural physical effects of a pumping turbine engine, the indications on the

[59]The phrasing could perhaps be made more correct from an English language point of view, but I have tried to translate the quoted Swedish phrasing as directly as possible.

instruments might not create an appropriate impression of urgency.[60] The spatial distance seems to inhibit or reduce natural feedback that probably is significant for an adequate understanding of system workings. In addition, the fact that airliners normally operate within safe limits and thus constitute a low-risk environment might eventually create a false sense of safety and a lowered preparedness for action. Combined with a lack of training for identifying and counteracting compressor stall and pumping, this might imply a contextual and emotional distance as well (R1.2.1). In terms of R1.1.5 (Figure 6.1, p. 117), the appearance of the situation to the pilots, endorsed by a lack of up-to-date skills and experience to identify and have confidence in the ability to handle such situations, did not trigger the will to make active use of available controls in order to control the system until it stopped pumping. In addition, there was this ATR system, adding to the difficulties for the pilots. In terms of R1.1.3 (ch. 5.4 – Character of computerized systems), the ATR is an example where at least two of the three characteristic properties of computerized technological systems implies undesired consequences. The system was working according to implemented models (C2) assuming that certain procedures were in use and that an immediate increase of thrust would be beneficial, which for this case was wrong in both aspects. It was also integrated (C3) covertly thereby inhibiting a natural system understanding for the pilots.

The edge awareness chart, R1.2.4 (Figure 9.3, p. 154), can be used to summarize all this. According to SHK, the pilots did not have sufficient situation awareness, in part due to inadequate training. However, they had, evidently, enough situation awareness to perform an emergency landing, although they were not enough aware of the situation for the engines, they did not address the pumping. From the present perspective, what they lacked was primarily sufficient system awareness, which SHK mentions only as an organizational problem, in the sense that ATR was unknown to SAS at the time of the accident. What seems to be overlooked is the generative consequence of the combination of these insufficient awarenesses, the fact that this made the pilots also lack sufficient edge awareness. They evidently fell over the operational edge for the engines, resulting in a complete breakdown of both, apparently because the pilots did not see the performance edge coming. They slipped, so to speak, on the ice that flew off the wings and thereby they ended up closer to the performance edge (the pumping edge) than ever before, without realizing it. Partly because of the

[60]Believe me, when having such an engine pumping in closer proximity, the identification of the situation is trivial and the urgency to act is more than obvious. I know this because the MD-80 series have engines from the same family as the Saab 37 Viggen fighters, although in the latter case optimized for the fighter application and fitted with an afterburner. When experiencing pumping in a Viggen, in which you sit slightly above and merely a couple of meters in front of the engine, the task to make the pumping stop does, so to speak, advertise itself, and you will indubitably attend to it with a very high priority (admittedly, not least because in the Viggen there is only one engine that keeps you flying...)

ATR, they did not have appropriate awareness about the performance relation between throttle settings and engine workings under such exceptional conditions with several compressor fan blades damaged, which would have been the necessary awareness for getting away from the pumping edge. This is, arguably, what caused the accident that was triggered by the presence of ice, a conclusion with an opposite ordering between aspects compared to the accident investigation report.

12.2 Forsmark, July 25th, 2006

The incident began at 13:20 on July 25th, 2006, when the Swedish national grid (SVK), the state-owned electricity-distribution organization, was about to do some work at the 400kV switchyard outside the Forsmark nuclear power station.[61] Forsmark power station consists of three boiling water reactors (BWR), where reactor number one and number two are connected to the switchyard in question, while number three is connected to a different one. Reactor number two was shut down for maintenance, while number one and number three were in full operation. In the switchyard to which the running Forsmark one was connected, a high-voltage disconnector was opened such that an arc appeared. This caused a two-phase short-circuit that in turn created severe fluctuations in voltage within the power station.

It is crucial for a nuclear power station to have access to electrical power, as this is required for maintaining control of the heat producing nuclear reaction process. Thus, ironically, for an electrical power plant, the availability of electrical power constitutes its weak spot. The power station at Forsmark have several sources of electrical power to guard this weakness, from the external 400kV power grid, from the external 70kV power grid, and from the in-house electrical power production. Furthermore, there are four independent internal power distribution subsystems, called subs, labeled A-D, where each one is fitted with a battery secured Uninterruptible Power Supply (UPS) system, designed to provide the station with electrical power for two hours, and one diesel driven generator per sub for prolonged emergency power. The battery system is connected to the sub between a rectifier and an inverter (because, batteries provide direct current and the

[61]This incident was not actually treated as an accident by the Swedish authorities, in the sense of being subject to an independent investigation. Information for the event description (ch. 12.2) and for report findings (ch. 12.2.1) are gathered from the publication called 'Bakgrund', which is Swedish for background (arguably meaning conditions or prerequisite knowledge), published by 'the analysis group' at the Nuclear Training and Security Center (KSU), which is a commercial organization thus not possible to be considered independent and unbiased. There are two versions of the text referred to here, one in Swedish (KSU 2006), and one in English (KSU 2007). Both have been consulted and they are collectively referred to simply as KSU.

power grid require alternating current). It is sufficient for two subs to function in order to provide the power station with necessary internal power.

What happened was that both circuit breakers connecting the main generators to the 400kV grid tripped on under-voltage, thus disconnecting the power station from the external 400kV power grid, resulting in a brief but substantial over-voltage on the internal electrical network. This triggered immediately a partial scram[62] and a switch was made to house-load production mode generating power only for the need of the power station itself. Two seconds after disconnecting from the external power grid two rectifiers tripped on a control fault and the inverters tripped on over-voltage, bringing down the UPS systems for subs A and B. That is, two out of four uninterruptible power supplies became interrupted. Fortunately, the rectifiers and inverters for the UPS systems at subs C and D did not trip. However, this made the station vulnerable as two working subs was stated as the absolute minimum. The manager initiated first incident response checks according to Emergency Operating Procedures (EOP).

After five seconds, one of the two main turbines made an emergency stop due to low governing oil pressure. Sub A was after eighteen seconds switched over to be directly supplied by the battery secured UPS due to low voltage, although the UPS was not working, which resulted in two seconds without power. This made the instrumentation chain powered by sub A shut down and consequently channel A of the emergency stop chain became triggered. The reactor power production was thereby reduced automatically to 25%, partly by a reduced flow of water through the plant, and partly by the insertion of some control rods as the result of the partial scram.

After twenty-four seconds sub A and C began to suffer from too low frequency and the normal supply circuit breakers opened, which made the instrumentation chain supplied by sub A once again lose its power. The diesels for sub A and C were started and connected, but it failed for sub A because switching it in required electrical power from the UPS system. The second main turbine made an emergency stop after thirty three seconds due to high pressure in the condenser. After thirty-five seconds, there was a changeover to a direct supply of sub A due to low voltage making the instrumentation chain supplied by sub B be without power for two seconds, and channel B of the emergency stop chain tripped resulting in a complete reactor scram because only two subs were working. After thirty-six seconds, one generator circuit breaker trips on low power and there was a changeover to the 70kV external grid for sub A and sub C. After thirty-seven seconds, sub B and D began to suffer from too low frequency just as A and C did, and the diesels were started and connected, but it failed for sub B as the UPS was

[62]Scram, or SCRAM, is a commonly used notion for a quick stop procedure at nuclear power plants. The origin of the word, or the acronym, if it actually is an acronym, appears however somewhat ambiguous and is therefore here assumed to be understood.

not working. When forty seconds had passed, the shift manager called for additional help from the plant specialists and the incoming afternoon shift.

Electrical power is required for equipment measuring water level and steam pressure within the reactor, and for control room instrumentation. Since not all subs were working and different equipment was connected to the different subs, some control room systems were not working, and the staff had to control the nuclear reactor in partial blindness.

After forty-three seconds, the second generator circuit breaker trips, resulting in no more in-house power being produced. There is a changeover to the 70kV external grid for subs B and D. When forty-five seconds had passed, the first checks according the EOP were carried out.

A scram is carried out by inserting all control rods into to the core, and when they are fully inserted this should be indicated in the control room. The insertion can be made in two ways, either quickly by use of a hydraulic system or a little slower by use of electrically driven screws. Because this was a full emergency scram, the hydraulic system was used. However, the signals indicating that the rods had been fully inserted should have come from the electrical screws of which only half was working due to subs A and C being down. Therefore, signals from neutron flux detectors in the core had to be used to assess whether the reactor output was as expected.

Systematic checks were carried out and when five minutes had passed a falling water level in the reactor pressure vessel was noted. After eight minutes, there was still no indication of all control rods being in place, although the neutron flux detectors indicated that the reactor was completely shut down. When fourteen minutes had passed it was noted that two out of four auxiliary feed water pumps were working, which actually was considered to provide sufficient cooling water flow to the reactor. However, the water level in the reactor pressure vessel was still falling and a specific check was made that at least two circuits in the emergency core cooling system were in operation. The first round of the EOP was then completed and the manager held a short meeting with the other operators.

After the meeting, when twenty-two minutes had passed, there was a successful manual restoration of power to the sub A and B diesel generators, resulting in all four subs having power again. The surveillance systems in the control room were restored and the remaining control rod screws were activated providing indication of all the rods being in place. Water pumping capacity increased again and normal levels could be maintained. Forty-five minutes from the beginning of the event, after extensive checks had been made, the staff was able to note in the log “The reactor is safely sub-critical and operational status is stable” (KSU, p.4).

12.2.1 Report findings

The incident was analyzed system by system, focusing on the different sub-system in sequence, the switchyard, the generator circuit breakers, the uninterruptible power supplies, the diesel generators, and the control room. This analysis was in the report then followed by one section labeled “What would have happened if...?”, another section labeled “Follow-up and lessons-learned”, and finally, “Conclusions”.

The switchyard short-circuit was apparently caused by a misjudgment by the external power grid organization, about the need to interlock an earth fault protection. Had this been done properly then there had been a much shorter short-circuit and much less severe fluctuations in voltage. If this had been the case, the disturbances had probably not affected the internal power production of the power station at all.

The generator circuit breakers for both main generators did not work properly. They should have opened on under-frequency when the generators were stopped, which they did not. In 2005, new under-frequency generator-protection systems had been installed, systems unknowingly working differently than the replaced ones. The old protector systems were independent of phase sequence in the three-phase grid while the new ones were not. Lacking knowledge about this fact made also the testing after the installation fail to identify the error. If the circuit breakers had worked as they were supposed to, then the power supply for switching in the diesel generators would also have worked.

Obviously, the UPS:es did not work according to expectations either, since two out of four failed to deliver power during this incident, which clearly is to fail being uninterruptible. The UPS:es were modern electronically enhanced systems installed according to supplier recommendations more than ten years earlier, replacing older systems based on more mechanical technology. The over-voltage protections were according to specifications designed to work within a voltage variation between 85% and 110% of the nominal value, and tests made by the supplier after the incident confirmed that they worked as expected. However, during this incident the voltage variations were significantly larger, meaning that from the perspective of the supplier the UPS:es were expected to fail. There were, however, small differences in the electrical circuits for the four UPS:es, which perhaps could explain why two of them sustained the voltage variations while two did not.

The diesel generators were automatically started, all four of them, but only two could be connected to their respective subs. Two of them failed because they required power from non-working UPS:es to establish the connections. The report concludes that this shows two things, it shows the vitality of UPS functionality, and it shows the fact that there were functional relations between the different systems, relations that made it possible for all to fail by a common cause.

Regarding the control room, the report does not say much. It highlights that the staff successfully carried out emergency procedures according to how similar incidents had been handled during earlier simulator training sessions, and, that despite a confusing situation with failing displays and lack of information, the staff managed to carry out "their work in accordance with their instructions in a particularly effective manner" (KSU, p. 5).

When answering questions about what would have happened if..., the report is arguably focusing on being reassuring to the public. It begins for example by establishing that the answers are provided by trustworthy expertise, presented jointly by the Forsmark Power Group (FKA) and the Swedish Nuclear Power Inspectorate (SKI). The section is rather short and because of its apparent bias to downplay the severity of the incident, it is recaptured here, completely as it is written in the (English version of the) report, following the initial statement.

> – If more than two UPS systems and the associated diesel generator units had not worked, there would still have been a good margin against boil dry of the core and damage to the fuel.
>
> – If three subs, rather than two, had been knocked out, the control room staff would have manually initiated forced blow down of steam from the reactor to the rector containment condensation pool.
>
> It would have been possible manually to energise the three diesel backed bus-bars within about 20 minutes, using power from the normal station distribution system.
>
> – If all four subs had been knocked out, it would have been possible to energize the diesel backed power systems manually from the normal power systems, with sufficient time to ensure good margins against core damage.
>
> – If none of the diesel generator units had been brought on line, and if additionally the 70kV grid had been without power, it would have been essential for the operators to act to obtain a power supply within 40-60 minutes through manual action. A limited amount of damage could then have occurred to the fuel.
>
> – If all four subs had been without power, *and* the operators had been unable to start correcting the situation within eight hours, it is very probable that serious damage to the core – a core meltdown – would have occurred. In this case, the various systems intended to limit the effects, in the form of the reactor containment's filtered pressure relief, would have come into play without requiring operator action. These systems would have prevented serious releases of radioactive substances to the surroundings.
>
> This is a situation which, in terms of effects on the surrounding area, can be compared with the reactor accident at Harrisburg in 1979. In spite of a reactor core melt down the releases of radioactive substances to the surroundings were small and negligible from a health point of view (KSU, p. 5).

The follow-up section concludes that there were significant shortcomings in barriers and in the defence-in-depth concept, the incident at Forsmark was categorized as belonging to category one, the most severe of the three categories for shortcomings defined by SKI. This means that there were

serious deficiencies detected in one or more barriers in the defence-in-depth approach, with justified suspicions that safety was seriously threatened. The mandatory action when a category one situation occurs is to immediately bring the plant to a safe state followed by a cold shut down, and before it may be restarted the results of the mandatory investigation must be reviewed and approved by SKI. The operator (FKA) did however decide to keep the reactor in a hot shut down state, which is put forth as the normal procedure after a scram to perform certain analyses. The cooling down of the reactor was not started before about 24 hours after the incident was initiated, which made SKI pass the case to the Prosecutor's Office at the end of January 2007.

Detailed analyses have also been presented to the International Atomic Energy Agency (IAEA) and the World Association of Nuclear Operators (WANO). On the IAEA scale for rector accidents, the INES scale starting at zero for a minor non-compliance to seven for a major accident (e.g., Chernobyl), the Forsmark incident was categorized as a level two. The idea with passing analyses on to IAEA is that other power stations should be able to draw experiences from this incident. One significant result of the analysis made after the incident, was the identification of a gradual deterioration of the safety culture within the company, over the last few years prior to this particular incident.

The conclusions are also, arguably, focusing on being reassuring. Although they do highlight the main flaw, that the designers of the defence-in-depth system assumed that voltage variation could not exceed certain limits thereby facilitating a single cause for multiple failures. The report concludes also that the training of operators in simulators proved to be valuable and kept control room staff working rationally in the stressing situation of this incident. It ends by concluding that the loss of power for the country as a whole, as the result of four nuclear reactors becoming shut down after the incident, went by without causing too many problems, a fact that probably was due to the warm weather and low need for electricity.

12.2.2 Interpretations according to present frameworks

To begin with, in the above recapitulation of the report, the initially stated underlying questioning of the analysis independence must shine through. This, for some perhaps slightly controversial, stance calls thereby for some additional motivation. The issuing organization, the nuclear training and safety center (KSU) was formed by the Swedish nuclear power companies in 1972, at the time when the first Swedish production-oriented nuclear power plant was started at Oskarshamn. KSU is today fully owned by Vattenfall, the state-owned electrical production and distribution company in Sweden. KSU is, among other things, responsible for operator training and competence development for nuclear power station workers, for this they use, among other things, simulators that are copies of the control rooms of

the power plants. The analysis group that actually wrote this report is stated as being an organization independent from but administrated by KSU.[63] These are facts that perhaps make the idea of analytical bias not that far-fetched after all. The in some sections unabashed downplay of severity and highlighting of successful aspects such as beneficial effects from the simulator training that they provide themselves is, to put it mildly, disturbing. All while this kind of confirmatory reasoning is, arguably, symptomatic within calculative safety cultures, which is a culture that appears to be the norm within the nuclear power production community. However, the analysis made by Forsmark themselves, identified what they earnestly labeled as a 'deteriorating safety culture' that in light of the accident report can be interpreted as a culture becoming too calculative.

The control of nuclear reactions is one of those things that human beings cannot do without help from technological systems. For such matters, technology is not merely enhancing but instead fundamental because it is what facilitates the very activity, making the design thoroughly influential on how the work will be performed. We human beings cannot observe radioactive emissions directly, nor withstand them for long. The processes are too quick, too hot, too powerful, and too dangerous for us to handle manually, and we are therefore confined to rely on feedback from artificial systems for judging situations and for selecting control input. The domain of effects (for the plant in general, but particularly for the core nuclear reaction process) is completely separated from the work-domain (R1.1.1 – ch. 5.2), especially from the perspective of control room operators. The central problem with all activities conducted from a distance is that available feedback can only be what is technologically feasible to convey and what in advance is assessed as enough valuable to convey that it merits costly development and implementation of instrumentation systems. This means that for any kind of unpredicted event there might be relevant observations for system operators that the technology fails to convey, simply because such an event had never been thought of before.

Problems related to consequences of the characteristic properties of computerized systems (R1.1.3 – ch. 5.4) are relevant for the breakdown of the two UPS:es. The UPS:es had built-in presuppositions (C2) about the stability of the external electrical environment. Their design created obscured and discrete performance edges (C1), edges that the systems apparently fell over without alerting the operators, and the UPS:es were also integrated in a way that made other parts of the defence-in-depth system fail as well (C3 – regarding basic system integration. Presumably, this particular integration did not have that much to do with abstracted and computerized interaction). It seems for this event, it was pure luck that the remaining two UPS:es did not fail as well.

[63]http://www.ksu.se/om-ksu – in Swedish, accessed 2013-06-04

The calculative approach to safety is salient throughout the analysis, but perhaps particularly in the 'what if...' and conclusions sections, where the reasoning centers on measures that will make the power station be safe (as a static state) if similar incidents would occur in the future. It is indicated, for example, by the focus on simulator training. Simulators are great for some purposes. One of their major flaws, however, is inherent in their very nature. They can only be used for training of predicted events because for valid training to be possible it must be made sure that the simulators behave reasonably similar to how the real systems would behave, which is possible only if the event can be analyzed in advance such that relevant system properties can be implemented in the simulator. This very incident was, as most incidents leading to accidents appear to be, an unpredicted event. It could not have been trained for in simulators because before the accident the common knowledge was that the UPS:es worked as they should, which means that the simulators would have been programmed such. If the malfunctioning had been predicted, the UPS:es would either have been modified or the staff would have been prepared to do the manual reconnection of the diesels directly when it occurred. This is the core paradox of calculative safety. By focusing on preventing undesired events from ever happening, that is, by considering the establishing of safeguards for predictable events sufficient, the safety becomes brittle because of an increased unpreparedness for unpredictable events, which in turn makes an unpredictable event more likely to cause a disaster. The report tries also to give the impression that the safety has been increased after the accident because this kind of event has now been implemented in the simulators, while this in reality only means that the safe-guard perhaps has become stronger, but probably also more brittle.

How to elude the calculative trap and the safety paradox seems a hard nut to crack, which perhaps can be illustrated by using the car driving metaphor once again. Car simulators and rally computer games can very well be used to learn the principles of car maneuvering and skid control, and they can also be used for memorizing track layouts and certain visual cues for known race tracks, something that probably will prove beneficial when a real racing event takes place on those tracks. However, the specific feel and dynamic workings of a specific and real car while skidding around the real track with its specific and current set of potholes and gravel ridges, these rich aspects cannot be trained in simulators. There are much too many essential but subtle nuances in the real situation, aspects impossible to convey in a simulator simply because the simulator is a map, not the real thing. To become a skillful rally driver it is therefore required to experience a lot of real world skidding in several and different real-world cars on multiple different real-world tracks. It is, arguably, only possible to master such a variety of subtle nuances by having a suitably well-developed intuition. In this case intuition is referring to system 1 (or system 1½) and the thinking

fast qualities of human beings (ch. 9.1), and not to some kind of supernatural sixth sense. Would it not be reassuring if the operators of nuclear power stations were as skillful in controlling their plants as rally drivers are at controlling their cars? This would however require the operators to practice maneuvering real plants under real circumstances in a way that is not possible. How could you practice skidding with a real nuclear power plant in the real world without risk causing terrible disasters? Allegedly, this was in some superficial sense what they tried to do at Chernobyl in 1986!

At present, it seems therefore for nuclear power plants that it is difficult, if not impossible, to get beyond the calculative level of safety. Consequently, if a new pothole happens to appear along the track, and they tend to appear regardless the preparation of the track, it is likely that the operators will be unable to handle a possibly resulting skid. Neither is it certain that an automatic safety system would manage, as this would require the event being predicted and a suitable automation developed in advance. Furthermore, the approach within the calculative vicious circle culture (R3.2) is to counteract this issue by broadening the road, put tarmac on it, and instruct operators never to leave the highway, which also provide beneficial grounds for further automation that, unsurprisingly, tend to work flawlessly under such suitably arranged conditions. This is what creates the brittleness. If the plant, due to some unanticipated combination of events, or perhaps because of a deliberate sabotage by evil terrorists, despite all thinkable precautions happen to stray down an unexplored alley eventually leading to a curvy gravel road with potholes. If that happens, highway bred drivers forced to use cruise control and steering aid systems optimized for long haul freeway driving will more than likely skid off the road.

In fact, an event that bears quite some resemblance with this incident from 2006 has happened again at Forsmark (at least superficially similar, there is no analysis such as the KSU report available yet).[64] Fortunately, it did not lead to anything severe this time either. On May 30th, 2013, in fact exactly while the first sections of this very chapter was being written, the third reactor at Forsmark, which was shut down for scheduled maintenance, lost its external electrical power. It began when maintenance work on one generator caused one of the three electrical phases in the power distribution to the plant to disappear. The reserve power did not connect automatically, forcing the personnel to start the diesel generators manually. The Swedish

[64]The information was gathered on the 7th of June 2013 from two news items published on the web. The first item is *Forsmarks reservkraft kopplade inte på* (eng., The backup power at Forsmark did not connect, my translation) by NyTeknik, a Swedish technology oriented weekly paper, http://www.nyteknik.se/nyheter/energi_miljo/karnkraft/article3707577.ece, referred to as NyT. The second item is *Forsmarks 3 förlorade tillfälligt elinmatning under revision* (eng. The 3rd reactor at Forsmark lost briefly its external electrical power during maintenance, my translation) issued by the Swedish Radiation Safety Authority (SSM), http://www.stralsakerhetsmyndigheten.se/Om-myndigheten/Aktuellt/Nyheter/Forsmarks-3-forlorade-tillfalligt-elinmatning-under-revision/ referred to as SSM

Radiation Safety Authority (SSM) rates the incident as a category one (the most severe classification) just as the 2006 incident, requiring a renewed clearance before the plant can be restarted.

One official at SSM says (to NyT) that the most severe aspect of the incident was that the plant did not handle the loss of adequate external power automatically, which is severe because that is what it is designed to do, adding that he cannot recall that this error has occurred before, neither at Forsmark nor anywhere else. Either he defines 'this error' narrowly enough to consider it a completely different error compared to the 2006 event, or he knows something significant not conveyed by these particular sources of information. At a general level, these two events appear extremely similar to me. They both started with an external disturbance in the electrical power distribution system causing the uninterruptible power supply system to be interrupted because automatic systems did not work as expected. The communication official at Forsmark says (also to NyT) that if the plant had been in production mode then this disturbance would have made the reserve generators connect automatically. The diesels did however not connect automatically because the second power connection to the plant was disconnected due to scheduled maintenance, he says. Is not the calculative reasoning painfully evident (things would have worked just fine if all systems had been working as expected...)? Moreover, was this not exactly the case for the 2006 incident (where the disturbance came from the external connection and the reserve generators did not connect because the UPS:es were down)? The calculative safety culture appears to remain.

12.3 Air France AF447, June 1st, 2009

Air France flight AF447, from Rio de Janeiro, bound for Paris, disappeared over the Atlantic on June 1st, 2009.[65] It seems that extensively automated subsystems and a highly abstracted control of the aircraft, combined with inadequate pilot training, played a significant part in this disaster, which should be considered a Wake-Up Call (Thomas 2011) to start learning the Painful Lessons (Flottau and Wall 2011) about consequences of too much automation and exaggerated reliance on predetermined behaviors.

Late Sunday night on May 31st, 2009, the Airbus A330-203 registered as F-GZCP was scheduled to leave Rio de Janeiro Galeão for Paris Charles de Gaulle, and it took off at 22:29 UTC.[66] About three hours later, around 01:35,

[65]The accident investigation report (BEA, Bureau d'Enquètes et d'Analyses pour la sécurité de l'aviation civile 2012) is in this chapter (12.3) referred to as BEA. The initial recapitulation of the event, and the recapitulation of the accident investigation findings, are made with the aim to convey the content of the BEA report without any significant alterations, but still modified into a hopefully more readable and less technical format.

[66]All times are Universal Time Coordinated (UTC), which at the actual time of the year means adding two hours to get Central European Time (CET) and subtracting three hours to get local

early on June 1st, at FL350 (about 10700 meters),[67] the co-pilot adjusted the level of detail on his navigation display and noted, "So we've got a thing straight ahead".[68] The Cockpit Voice Recorder (CVR) system was running and conversations as well as radio traffic was continuously captured. This 'thing' that the co-pilot referred to was a bit of bad weather related to the Intertropical Convergence Zone (ITCZ),[69] something the crew also had been informed about by the Operations Control Center (OCC) around one hour earlier. The Captain confirmed the statement and they began once again discussing the fact that they still had a very heavy aircraft[70] and that the comparatively high temperature inhibited them from climbing to FL370 (about 11300 meters) for an attempt to get above the bad weather. They dimmed the internal lights to be able to see better out through the cockpit windshield and the co-pilot observed that they were "entering the cloud layer". Turbulence started shortly after.

Flying fast at high altitudes implies a delicate compromise between several aspects, especially for airliners that are subsonic aircraft (with the now retired Concorde as the obvious exception). To begin with, the angle-of-attack[71] required for maintaining the flight-level increases at high altitudes because of reduced air pressure and aerodynamic speed, and here it was further increased by a comparatively high temperature. At the same time, the angle-of-attack where the wings begin to stall[72] and lose their lift-generating capacity is decreasing with increasing Mach numbers.[73] Moreover, higher altitude can mean lower fuel consumption per ground distance due to reduced drag, less oxygen, and higher relative ground speed, but it means also less available thrust, in a situation when more thrust is required because

---

time in Rio de Janeiro.

[67]Flight Level (FL) is the standardized altitude reference system used for separation of aircraft cruising at high altitudes where the exact height above the ground is irrelevant. It is given in hundreds of feet according to the standard sea-level atmospheric pressure (1013,25hPa)

[68]All unreferenced quotes are from the report, such as all crew statements from the CVR. The report's annexe.01 (CVR transcript) and annexe.02 (FDR [flight data recorder] chronology) have been extensively consulted, and inconsistencies have been resolved by giving the annexes higher priority than the main report (there were actually several inconsistencies identified, especially among CVR quotations).

[69]The ITCZ is a zone where the northeast and southeast trade winds converge, its position varies but is usually located near the equator, especially over the ocean. This convergence results in several weather phenomena, where heavy precipitation and large cloud formations like cumulonimbus (thunderstorms) are frequent.

[70]Long-haul airliners take off with extensive amounts of fuel, making them overweight and subject to extra restrictions in maneuvering, until some fuel have been consumed.

[71]Angle-of-attack is the angle between the incoming airflow and the wings. A larger angle (i.e., nose-up) means higher aerodynamic lift, until the wing stalls. At ground level and at optimum speed a typical general aviation wing has a maximum angle-of-attack of about 20°

[72]Stall as a phenomenon was elaborated on in footnote 10, p. 25.

[73]Mach number is the dimensionless ratio between the current speed and the speed of sound, which is where the aerodynamic properties of the air shifts significantly in character. The speed of sound depends mainly on temperature, and it is reduced along with the falling temperature at increasing altitudes.

the angle-of-attack is increased, which in turn implies higher fuel consumption. Altogether, this is not an easy puzzle to sort out, and for this particular case, the important thing to note is the diminishing stall-margin at high altitudes (i.e., the difference between maximum and required angle-of-attack).[74] To assist in this matter the Flight Management System (FMS) of the aircraft in question calculates two altitude recommendations, one labeled OPTI that takes cost parameters such as ground speed and fuel consumption into account, and one labeled REC MAX considering only aerodynamic aspects. REC MAX is always higher than OPTI but lower than the certified maximum altitude of the aircraft. Air France crews were used to keep some margin to indicated REC MAX.

The turbulence stopped shortly after 01:52 and the co-pilot once again drew the attention of the Captain to the REC MAX value, now indicating FL375. The Captain did however choose, as was customary to do about that time of flight, to wake up the second co-pilot, saying “well right he's going to take my place”, and directly after, “You're a PL aren't you?”, meaning to confirm that the co-pilot in fact was approved to act as relief-Captain, which was acknowledged by the co-pilot. The Captain stayed in the cockpit until around 02:00 and attended the briefing between the two co-pilots. Where the Pilot-Flying (PF) seated on the right side, explicitly pointed out to the newly arrived Pilot-Not-Flying (PNF) now seated in the captain's place on the left side, that there was some bad weather ahead, which they were unable to try getting above due to the high temperature. He also informed about their failure to log on to DAKAR (the Oceanic Area Control Center). The Captain aged 58 and overall the most experienced one of the three pilots, went for in-flight rest, leaving the PF aged 32 and least experienced as the only one fully updated on the course of events for this flight so far, yet designated relief-Captain. The newly arrived PNF, aged 37, was senior to the PF, actually three times more experienced on this particular aircraft type, with South American trips, and with ITCZ crossings, even compared to the Captain himself, and as appointed cadre at the Technical Flight Crew Division the PNF was “enjoying recognition as an expert by his peers” (BEA, p. 170). While there was no sign of conflict or confusion of roles between the two co-pilots, the hierarchical structure, formal and informal, was in fact slightly unclear, something the Captain overlooked or ignored to address.

The two pilots remaining in cockpit discussed once again the conditions for REC MAX while the turbulence began to increase around 02:06. The PF informed the cabin crew that “in two minutes there we ought to be in an area where it will start moving about a bit more than now you'll have to watch out there”, adding “I'll call you when we're out of it”. Two minutes later the PNF asked the PF “Don't you maybe want to go to the left a bit?”, followed by a

[74]This delicate part of the flight-envelope (the operational limits for the aircraft) is sometimes called 'the coffin corner', e.g., http://en.wikipedia.org/wiki/Coffin_corner_(aviation)

few thoughts about how to diverge from the planned route while being in autopilot mode, the speed was reduced and engine de-icing was turned on.

At 02h:09m:40s, there was a change in the background noise, later identified by A330-340 pilots as the sound of ice-crystals hitting the aircraft. The two pilots discussed whether one of them had done something with the climate settings, although it shifted quickly into a discussion about the smell of ozone that had appeared. After twenty-five seconds, at 02h:10m:05s, first the autopilot then the auto-thrust systems disconnected and the PF called out "I have the controls". The aircraft began to roll rather quickly to the right and the PF made a distinct nose-up and left input, followed by two left-right inputs to the stop positions. Supposedly, the excessive control inputs came from being surprised by the quick initial right roll combined with an unfamiliarity with and unawareness about the change into alternate flight control law.[75] The roll angle fluctuated between 11° right and 6° left while the pitch attitude[76] increased to 11° in ten seconds. The indicated speed on the left Primary Flight Display (PFD)[77] made a sharp fall from about 275kt to 60kt and shortly later on the Integrated Standby Instrument System (ISIS) as well, and there were two brief stall-warnings. The Flight Director (FD) indications on the PFD disappeared without the crew explicitly disconnecting them, indicating the loss of normal flight control protections.

The PF said at 02h:10:15s "We haven't got a good display … of speed", the PNF said simultaneously "We've lost the speeds so … engine thrust A T H R engine lever thrust", which came from reading out messages on the ECAM-display (Electronic Centralized Aircraft Monitoring), and six seconds later he read out "alternate law protections...". The PNF also called out that he turned on the wing anti-icing, which also became turned on. This division of tasks was completely consistent with expected procedure, where the PF is supposed to focus on handling the aircraft while the PNF should work on identifying and correcting the problem. From around 02h:10:25s and for about ten seconds, the PNF apparently became concerned about the aircraft handling because he repeatedly asked the PF to "Watch your speed", saying "we're going up", and "go back down", "gently". The aircraft was then at about 37000ft (~11280m), still climbing quickly at almost 7000ft/min (~35m/s) and the PF said "Okay, okay okay I'm going back down" and made several nose-down inputs that reduced the pitch attitude and climb rate for a while. The PNF continued "Stabilize", "Go back down", "According to that we're going up", "According to all three you're going up so go back down", "You're at...", "Go back down", "gently". The PF came in once in a while with "yeah", "okay", "It's going we're going (back) down",[78] but shortly after

[75]Flight control law and flight protections are notions describing the level of abstraction in control of the aircraft, these levels are discussed more in the following sections.
[76]Pitch attitude is the longitudinal orientation relative the ground, positive means nose up.
[77]The speed on the right display is not recorded by the Flight Data Recorder (FDR)
[78]Parenthesis within CVR transcript quotes indicates uncertainty about what actually was said.

“We're in … yeah we're in climb”. The PNF then turned his attention back to try getting hold of the Captain, supposedly as the result of remembering that his primary task was to solve the problem and not to fly the aircraft.

Between 02h:10m:36s and 02h:10m:50s, the speed on the left side became valid again while the ISIS continued to be erroneous. Thrust controls were pulled back slightly ending around 85%. The pitch attitude came back above 6° and the angle-of-attack was slightly less than 5°, but the pitch increased again progressively beyond 10° and the aircraft continued to climb. Then the stall warning triggered, and stayed on. The thrust was increased to maximum (TO/GA)[79] while the PF made nose-up inputs resulting in the increasing angle-of-attack beginning at around 6° when the stall-warning started. There was also the Trimmable Horizontal Stabilizer (THS) system, which added to the upward strive by moving from 3° to 13° pitch-up and remaining there until end of flight. Around 02:11, the third Air Data Reference (ADR3) was selected on the right FDR and at the same time the ISIS became valid. The three speed displays were now consistent showing 185kt and the aircraft peaked at its altitude at about 38000ft (~11600m). The PNF, apparently beginning to worry again about the handling of the aircraft, said “Above all try to touch the lateral controls as little as possible eh”. Then he asked openly “But we've got the engines what's happening … ?”, and shortly later he asked the PF explicitly “Do you understand what's happening or not”

The PF complained at 02h:11:32s that “I don't have control of the airplane any more now”, and again two seconds later “I don't have control of the airplane at all”. At 02h:11m:37s, the PNF said, “controls to the left” and took over priority, although the PF took back priority almost instantly, and did so without saying anything. The PNF asked “(...) what is that?”, and the PF answered, what perhaps could explain his continued nose-up control input, that “I have the the impression (we have) the speed”.

The Captain came back into the cockpit at 2h:11m:42s saying “Er what are you (doing)?” The PNF answered, “What's happening? I don't know what's happening” and the PF said, “We're losing control of the airplane there”. The PNF again, “We lost all control of the aeroplane we don't understand anything we've tried everything”. At that time, all recorded speeds became invalid and the stall warning ended. However, the angle-of-attack was now more than 40° and the plane was falling at about 10000ft/min (~51m/s), while the pitch attitude stayed above the horizon but below 15°. Thrust was reduced to minimum (IDLE). The PF made side-stick input to the left stop and nose-up for about 30 seconds. At 02h:11m:59s, he said “I have a problem it's that I don't have vertical speed indication”. Two seconds later, the Captain answered “alright” at the same time as the PF said “I have no more displays” and the PNF repeated, “we have no valid displays”, and once again the PF expressed his confusion about the speed by

[79]Take-Off/Go-Around

saying “I have the impression that we have some crazy speed no what do you think?” He apparently reached out to apply the air brakes, because the PNF said “no”, “No above all don't extend (the)” and the PF answered “no?”, “okay”, although the air brakes became deployed anyway.

None of the pilots seems to have realized that they were in a fully developed stall because the confused statements continued. The PF said at 02h:12m:11s, “So we're still going down” and twenty seconds later “Am I going down now?” followed by “I'm climbing okay so we're going down”, and further 15 seconds later “Yeah yeah yeah I'm going down, no?” The PNF said at 02h:12m:13s, “We're pulling” (supposedly meaning trying to pull up) and asked the Captain “What do you think about it what do you think what do we need to do?” and the Captain, apparently also confused, answered “There I don't know there it's going down”. At 02h:12m:27s, the PNF said “You're climbing”, but directly after he seems to have realized something because then he said “You're going down down down”. The Captain had obviously not yet arrived at the same conclusion because at 02h:12m:32s he said “No you climb there” followed two seconds later by “You're climbing”. However, it seems the Captain also began to realize what was happening, because at 02h:12h:44h he said “(...) it's impossible”, followed eight seconds later by “Hey you”, “You're in...”, “Get the wings horizontal” At that time the aircraft had already stalled down to about 20000ft (~6000m).

The PNF tried to begin piloting, while the PF also continued struggling with the controls. Just after the PF had concluded “We're there we're there we're passing level one hundred” (10000ft, ~3000m), the PNF said at 02h:13m:20s, “Wait me I have I have the controls eh”, but he seems to have let go and asked instead “Try to find out what you can do with your controls up there”. The PNF had after all apparently still not fully understood the situation, because when the PF at 02h:13m:36s called out “Nine thousand feet”, the PNF responded with “Climb climb climb”. Then the PF said what probably made both the Captain and the PNF realize what was going on.

At 02h:13m:41s, the PF said “But I've been at maxi nose-up for a while”. The Captain responded “no no no don't climb” and the PNF “so go down”, “So give me the controls the controls to me controls to me”, and the PF acknowledged “Go ahead you have the controls we are still in TOGA eh”, and the Captain continued “(so wait) AP OFF” meaning to shut off any remaining autopilot involvement. However, it was already too late.

The PNF side-stick was positioned nose-down for 15 consecutive seconds, supposedly in a futile attempt to make a stall recovery maneuver, although the DUAL INPUT parameter in the FDR was activated five times. At 02h:14m:05s, the Captain warned “Watch out you're pitching up there”, the PNF answered “I'm pitching up” having his side-stick positioned slightly nose-up, and the PF answered “Well we need to we are at four thousand feet” (~1200m). Then the ground-collision warning system triggered with a voice repeating “pull up”, “pull up”. At 02h:14m:21s, the PF takes control

priority and says “(!) we're going to crash”, “This can't be true”, and finally he uttered a last expression of confusion, “But what's happening”. The PNF side-stick is positioned nose-down, the PF nose-up to the stop. Then the recording stopped, at 02h:14m:28s.

12.3.1 Accident investigation findings

Just as for the SAS SK751 incident, this accident was also triggered by the presence of ice, in a situation where ice actually was known to exist and for which there were procedures and regulations in effect supposed to remedy the problem. While the ice in the SK751 case actually damaged the aircraft in a way that eventually led to the destruction of both engines and to a forced emergency landing (a crash), the ice-crystals hitting AF447 caused nothing but a temporary loss of airspeed. A loss that would have been completely insignificant and harmless for the aircraft, if not the situation had evolved as it did. This means that the crucial question is why did this, in some sense trivial, malfunction evolve into a fully developed stall right into the ocean.

The report begins with the fact that the crew became completely surprised by what happened and that they seem never have understood fully the situation they ended up in. For this, there may have been several reasons. To begin with, when cruising at high altitudes on long-haul flights the main concern for the crew is usually to avoid turbulence, for comfort reasons (BEA, p.168). This task involves selection of alternative flight paths and flight levels weighted against extra fuel consumption and prolonged flight time, a task that is performed on a rather high level of abstraction mainly by adjusting autopilot settings. The step to begin considering existential maneuvering of the aircraft by use of direct control input to keep it flying is thereby quite long, thus supposedly requiring significant indications to trigger the insight necessary for actually taking the big step mentally. The problems associated with identifying the situation correctly is thus the focus of the report.

If a total loss of airspeed happens to occur, this implies, obviously, a risk that the aircraft will reach either a too low or a too high speed without the crew noticing, and both situations are thoroughly undesirable. In addition, all the different autopilot systems lose necessary input and can no longer function properly. The autopilots are therefore designed to disconnect when required input is missing, or deemed ambiguous by certain implemented criteria, and since modern aircraft in normal operation rely heavily on autopilots a disconnect calls for a significant warning to be issued. It seems the crew mainly became concerned with the autopilot disconnection and failed to consider the reason why it did so, perhaps because there was no salient warning about the loss of airspeed (BEA, p.172). This inability to identify the underlying problem may also have been influenced by the fact that when the pilots were training for this kind of situations in the simulators,

the problem was presented solely as a loss of airspeed indication, with no other errors displayed (BEA, p. 175). In the case of this accident, the ECAM displayed several messages read out in an unstructured order by the PNF, which probably added to the inability to identify the loss of airspeed as the culprit and instead drew focus to the possibility of a more complex problem (BEA, p. 176). Furthermore, it seems the PF was mainly concerned with the risk for over-speed, and he apparently connected the buffeting that occurred with high-speed stall while failing to consider the facts of "the coffin corner"[80], where essentially the limits for high-speed stall and low-speed stall converge. The report concludes that system feedback about speed anomaly was poor and that the available information in fact was likely to create the impression of over-speed being the major risk (BEA, p. 174).

Another issue in the development of this disaster was the inappropriately excessive control inputs made especially by the PF, not at all in accordance with what is expected for a crew flying at high altitudes. To begin with, this seems to have added to the confusion of roles. While the PNF at large followed the standard procedure where the PF should focus on flying the aircraft and the PNF on identifying the problem, the PNF evidently became concerned about the handling of the aircraft. The side-stick design added significantly to this confusion, simply because it is virtually impossible for the one not flying to observe control input from the other pilot, and by that judge whether the situation is under control (BEA, p. 174). At one time the PNF asked to have the controls and was granted so by the PF who, however, almost immediately regained priority control without saying so, probably adding to the confusion for the PNF about the situation because his control inputs did not seem to make any difference.[81]

Regarding identification of stall, and stall warnings produced by the system, observations from other incidents with aircraft from the same aircraft family conclude that stall warnings are regularly considered surprising and irrelevant (BEA, p. 174), which supposedly also was the case in this accident because neither of the pilots made any comments related to the stall warnings. At the same time, the operational safety model and the listed emergency procedures expect pilots to identify the approach to stall in the order of a few seconds. However, crews practice this kind of identification solely during initial training and they are unlikely to encounter approaches to stall more than a few times during a career, and they are even less likely ever to have to deal with a fully developed stall. Yet the safety model assumes that their abilities will remain sufficient over time.

[80]The 'coffin corner' is the end of the flight envelope where the airframe can be at the edges of high speed and low speed stall simultaneously (cf. footnote 74, p. 197).

[81]This confusion would never occur in a classical yoked stick design where the controls are mechanically connected, thereby making the pilots see what the other one is doing and implying that muscular force is required in order to override each other's input.

The crucial moment of the event was the exit of the flight envelope, up until which appropriate control input probably would have regained safe flight rather quickly. This exit happened around the time when the aircraft peaked in altitude at about 38000ft (~11600m). After exiting the flight envelope, only very deliberate and consistent control input from an extremely resourceful crew could have saved the situation. This is because the only way out from the fully developed stall would have been to reduce the angle of attack significantly and gain speed by making the aircraft enter into a steep dive. For everyone not familiar with advanced aircraft handling and aerobatics (for which an airliner is not built) this would perhaps appear as worsening the situation, yet it had to be done for any chance of getting back within the flight envelope. However, this maneuver would obviously require quite a nerve and total insight by the crew of being in stall, together with sufficient altitude, which they actually had, although not for long. Hence, the fundamental problem seems to have been lack of situation insight.

The main problem was, apparently, that the crew never realized, at least not in time, that they were stalling. The PF's focus on the risk for over-speed and high-speed stall was probably also significant, and his failure to consider the risk of low-speed stall might have come from embracing the common belief that the airplane could not (low-speed) stall (BEA, p. 180). He was then rejecting the implications of alternate flight control laws and the fact that the flight control protections, as they were called, were not working, which perhaps could explain the excessive control inputs. It is also possible that he failed to realize the possibility for the Flight Director (FD) information to be invalid, all while the emergency response procedure for loss of airspeed says to disable the FD in order to prevent irrelevant cues (BEA, p. 181). This kind of deliberate action to disconnect what under normal conditions is a highly utilized assisting system presupposes, of course, that the loss of airspeed immediately is identified and comprehended, which appears not have been the case. The gradual deterioration of air-speed and altitude information may have produced FD indications to climb, which the PF might have responded to by reflex in the stressful situation.

There was also a crew-coordination problem. The “willing but chaotic cooperation in managing the surprise generated by the autopilot disconnection led quickly to the loss of cognitive control of the situation, and subsequently to the loss of physical control of the aeroplane” (BEA, p. 184). The fact that the captain failed to address the discrepancy between the informal and formal authoritative levels, amount of experience, and assumed expertise, of the two co-pilots may have contributed to the event. By designating the least experienced pilot as relief-captain may have hampered the more experienced PNF in sorting out the situation, and this problem is addressed in recommendations following the final conclusions.

The report concludes that what caused the accident was a series of events, beginning with the obstruction of the pitot probes by ice crystals (BEA, p. 199). This obstruction caused a total loss of airspeed and the autopilot to disconnect, and a reconfiguration of flight mode to alternate law. The crew failed to link the loss of airspeed to appropriate procedures and they failed to identify in time the deviation from the flight path and the approach to stall (i.e., they failed to comprehend the effects of next statement). Inappropriate control inputs made the aircraft exit its flight envelope and enter into a fully developed stall, which continued until even such inputs that could have made it possible to recover into safe flight were too late.

Following the list of events that were concluded to be the cause of the accident, the report lists a set of factors that in combination can explain these events. To begin with there were "feedback mechanisms for all those involved that made it impossible" (BEA, p. 200, here interpreted as a culture making it impossible) to identify and remedy the tendency not to apply the loss of airspeed emergency procedures when obstruction of pitot tubes occurred, nor to ensure that crews fully understand the consequences of pitot tube icing. The problem was known to exist, but seems to have been misunderstood by the aviation community (BEA, p. 199). Another aspect mentioned was the total absence of training in manual handling of the aircraft at high altitudes and of stall recovery procedures. The problems with cockpit task sharing was mentioned, presumably this was caused by the incomprehension of the situation when the autopilot disconnected, and from poor management of the emotionally charged situation. Consequences of aircraft design is mentioned from two aspects, first as the lack of clear indications of airspeed inconsistencies identified by the computers, and then as factors making the crew fail to take the stall warnings into account. The latter with a list of possible explanations: low exposure to the different stall phenomena during training, transient and apparently spurious warnings at the on-set of the event, absence of approach-to-stall information after losing the airspeed limits, confusion with over-speed situation, FD indications, failure to recognize and understand implications of alternate flight mode laws lacking protections.

12.3.2 Interpretations according to present frameworks

The Airbus A330-200 is a modern airliner from the A320 family, which introduced digital fly-by-wire[82] flight control technology and side-stick maneuvering for pilots in non-military aircraft designs. Traditional aircraft designs include mechanically coupled sticks or yokes (the commonly W-

[82]The term 'fly-by-wire' is actually somewhat ambiguous because in many "mechanical" aircraft the stick or yoke and pedals are connected to the aerodynamic control surfaces (i.e., ailerons, elevators and rudder) via steel wires. Fly-by-wire does however mean flying by use of electrical wiring, and implicitly through computerized flight-control systems.

shaped “steering wheels” of many aircraft) and pedals, also more or less mechanically connected directly to ailerons, elevators and rudder. The A320 had its first flight in 1987 and the A330 came into service in 1994. Military fighter aircraft had already been utilizing electronic flight control technologies for a while, for reasons quite understandable in military contexts. For example, the General Dynamics (now Lockheed-Martin) F-16 Fighting Falcon that had its first flight in 1974 and was introduced into military service in 1978. The F-16 is allegedly the first modern airframe deliberately designed with a relaxed static longitudinal stability (pitch instability). This design implies an increased maneuverability, which obviously is a desirable quality for a fighter aircraft. The delicate matter of controlling the unstable airframe, a task simply not manageable for human beings because of too quick dynamics, was therefore something that just had to be conquered by a technological design. Unstable airframes require control abstraction through a fly-by-wire design, which as a convenient side effect allows for cockpit designs freed from several mechanical constraints. For the F-16 this led to the characteristic and now legendary side-stick and bubble canopy design, providing extra dashboard space and very good visibility in most directions, for fighter aircraft another highly desirable quality because a good outside view is obviously quite beneficial when being in air battle.

Airliners do not require the kind of maneuverability that comes from instability. They require, on the contrary, a high level of stability, for comfort and safety reasons. To increase the outside view by using freed dashboard space seems neither be a top priority when designing airliners. This is in part because commercial flights to an increasing extent are taking place in highly regulated airspace, thereby following strict procedures supervised by air traffic controllers, implying an airspace situation rather having a decreased need for good visibility. Although, visibility is naturally important also for airliners, not least for a basic level of see-and-avoid capability, but the point is that outside visibility cannot have been a driving matter for the implementation of fly-by-wire and side-stick maneuvering in the A320 family. Airliners simply do not require unstable aerodynamics and abstracted fly-by-wire control. The design goals for introducing fly-by-wire in the A320 must therefore have been different, presumably economic and perhaps ergonomic, or was it mainly to show off with a new and modern design inspired by legendary fighter aircraft, not really considering operational implications?

Mechanical constructions tend to imply inflexible design solutions demanding both muscular strength and sometimes difficult-to-acquire maneuvering skills and short service intervals because mechanical systems are bulky, heavy, may be tricky to master, and they are maintenance demanding. Electronic systems, on the other hand, can be made small and light, they can be fitted principally wherever there is enough space because

electrical wiring is more flexible than mechanical couplings, and electronics often requires less maintenance. They can also be provided with arbitrary user interfaces that ideally are easy to learn. Furthermore, another aspect of increased computerization of airframe flight control, and of similar subsystems thoroughly grounded in physical matters like engine control, climate systems control, and such like, is the increased opportunities for, and easiness to automate. To integrate another level of functionality for a speed regulator or an autopilot system in a mechanically controlled aircraft require physical actuators and logic-mechanic interfaces such as hydraulics and electric servomotor mechanisms, which are systems that consume precious resources such as power and space thereby putting constraints on other parts of the aircraft design. To create the same kind of extended functionality with a computerized fly-by-wire system already in place simply means to implement suitable logics in software and tie it in at an appropriate level in the digital flight control system, which makes system integration significantly simpler (R1.1.3 – C3) and thereby cheaper. That is, the fly-by-wire design and the computerized control paradigm might primarily have been driven by economic aspects, rather than by desired system handling and performance aspects, or it may have been driven by the vicious circle culture (R3) confusing predetermined effects with situated desired effects.

From the perspective of the user, the result of computerized abstracted control is that the pilots do not need to bother with physical throttle levels and skillful direct control of ailerons, elevators and rudder. They can focus on higher order decisions such as desired climb or descend rate, heading, and speed, which for the Airbus A330 family of aircraft is the normal flight control mode. The level of control-abstraction is however a completely arbitrary design decision (R1.1.3) that in theory (although in practice for UAV:s) can end with a fully automated system in principle requiring nothing but a single triggering control input (R1.1.5, R1.1.2). Consequently, the actual flying of the aircraft has become abstracted to imply instead a controlling of the flight-control system, in this case also by use of a side-mounted flight-control stick. This normal flight control mode implies that control input is filtered through the computerized autopilot system (R1.1.1, R1.1.3, R1.2.1-U6, and R1.2.2) that has a number of built-in flight control protections that prevents the aircraft from leaving its flight envelope. That is, if the pilot gives nose-up input (i.e., increases the climb rate) more than the current speed and available thrust allows for, the flight control protections will stop this at the edge of the calculated flight envelope. This is presumably the basis for the stated and commonly adopted belief that the aircraft could not stall.

However, these control abstractions and protections require air data input, without which the computers cannot calculate the limits of the flight envelope. Air data is collected by the pitot tubes with apertures that if blocked by something, for example ice, stop working. If this would happen,

or if any other data inconsistency is identified, the flight control system is designed with a number of alternate flight control modes, of which the most basic one is called direct flight control mode. As the name suggests, direct flight control mode implies that the pilots control the ailerons and elevators and rudder directly, in principle as in a traditional and more mechanical aircraft, although still via electrical wiring and by use of the side-stick. While the benefits of this kind of design are many, there is a risk of losing touch with the basics, in this case with fundamental maneuvering of the real physical airframe in the real physical world. The joystick-like side-stick adds arguably to the detachment from the basics, as it can easily create an impression of controlling a computer-game (R1.2.1, in particular U6.3 and U6.4). The character of the side-stick handling made by the PF suggests also that he never understood that the aircraft had shifted into direct control mode, or at least he seems not to have understood its implications.

The report concludes that the cause of the accident was a series of events, starting with ice obstructing the pitot tubes, forcing a decoupling of autopilot functions and entering into alternate flight control mode. This was followed by a deviation from the flightpath and inappropriate control input that made the aircraft exit its flight envelope, and, because of the absence of appropriate recovery input, this resulted in a stall that continued all the way down into the ocean. This is a conclusion focusing solely on concrete actions and courses of events, while forgetting to consider what caused the conditions of this situation to be what they were. This conclusion is arguably a too narrow perspective implicitly demanding responsibility from people not given proper means to shoulder those demands.

However, the report provides some explanations that touch upon these matters. For example, the identification of a culture that made it impossible for people involved to address the implications of certain technological insufficiencies, such as the consequences of a total loss of air speed, by use of further technological solutions. This culture can also be seen as a part in the conclusion that pilot training in basic aircraft handling (at high altitudes) and stall recovery was insufficient. As already have been said, in terms of the present frameworks, especially R3, this culture may be equaled with a calculative culture. It is a culture considering such training and addressing of technological insufficiencies unnecessary because the technology is considered to make sure it is unnecessary, which it would be if reality would be equally predictable as the models governing the design of the technology.

However, it is possible to view the matter from the opposite viewpoint compared to the accident investigation report. The explanations, and especially the calculative culture, is what have created technological conditions in the form of system designs that made the situation evolve as it did, which thereby must be concluded as the cause of the accident. The series of events in the report stated as the cause is here argued better considered an explanation of how it evolved. The calculative culture

identified within the airline operations community seems unfortunately to extend to and prevail also in the airliner constructors community, resulting in aircraft designs with such an abstracted control that it effectively inhibits the pilots to do what they are supposed to do, namely to fly the plane. In terms of R1.2.4, it is possible to state that the approach during aircraft design and the resulting character of controllability (R1.2.2) led to insufficient system awareness, indicated by surprise, confusion, and excessive control inputs. The aircraft design led also to insufficient situation awareness, indicated by the failure of the crew to identify the stall and apply appropriate control input to recover. Consequently, the aircraft design led also to insufficient edge awareness, proved by the fact that they fell over the stall edge and exited the flight envelope without providing appropriate control input to remedy the situation. The vicious circle culture (R3), considering predictable behavior (R3.2) sufficient enough to be enforced by the abstracted control model (R3.3), had made the pilots adopt too much the role of being automation supervisors, at the expense of skilled system handling on the more fundamental level of aircraft aerodynamics.

12.4 Fukushima, March 11th, 2011

The Fukushima Daiichi Nuclear Power Plant operated by Tokyo Electric Power Company (TEPCO) was severely damaged by the Great Eastern Earthquake and the following tsunami.[83] These natural disasters triggered a man-made nuclear accident that is still happening because consequences are not restored and people are still suffering from exposure of radiation. The accident was man-made because it could and should have been foreseen and prevented. In fact, it was principally foreseen, but identified necessary measures were either ignored or postponed, and apparently, this was mainly because of the Japanese management culture. Kiyoshi Kurokawa, the chair of the National Independent Investigation Commission (NAIIC), put forth in his message at the very beginning of the report that:

> What must be admitted – very painfully – is that this was a disaster "Made in Japan". Its fundamental causes are to be found in the ingrained conventions of Japanese culture: our reflexive obedience; our reluctance to question authority; our devotion to 'sticking with the program'; our groupism; and our insularity. (NAIIC-Exec, p. 9)

[83]The official accident report (NAIIC 2012) consists of eleven documents plus an executive summary originally written in Japanese. The present work is based on the provided direct English translations. These documents are here collectively referred to by NAIIC and general statements put forth are attributed to 'the Commission'. This chapter (12.4), and ch 12.4.1 are aimed at being undistorted recapitulations of the report contents, therefore mostly lacking general references. Explicit quotes are however referenced, with a presumably identifiable shorthand for the different documents, e.g., NAIIC-Exec for the Executive Summary, NAIIC-Intro for the introductory chapter, NAIIC-Ch1, and so forth.

On March 11th, 2011 at 14:46, the Great Eastern Earthquake occurred, immediately triggering the SCRAM, the emergency shutdown feature, on unit 1, 2, and 3 of the Fukushima Daiichi Nuclear Power Plant. Units 4 to 6 were already shut down due to periodical inspections. However, the seismic tremors had damaged the electric transmission to the power plant, with a total loss of off-site electricity as the result. Even the secondary back-up line was unusable due to mismatched sockets. The tsunami that followed from the earthquake had its peak at 15:37. It destroyed the emergency diesel generators, the seawater cooling pumps, electrical wiring, and the DC power supply for units 1, 2, and 4. There was simply no electrical power available, except from an air-cooled emergency diesel generator at unit 6. Unit 3 had initially some DC power that however ended before dawn of March 13th.

The power supply was not the only thing damaged by the tsunami. Vehicles and heavy machinery was washed away, buildings were destroyed, oil tanks and lose objects were flushed around together with large amounts of gravel. The water reached as high as to the high pressure sections of units 3 and 4 and it flooded the radiation waste storage facility called the common pool building. When the water retracted, debris was scattered over the entire site, manhole and ditch covers were gone leaving trapping holes and obstacles that made movement on the site extremely difficult. Buildings were shifted and interiors were collapsed, and on top of this, there were several after-shocks from the earthquake. This made it impossible to reach the site with additional fire trucks and generator trucks as alternative ways of cooling the reactors and providing electrical power. Only one fire truck was available at unit 1 as the second truck was broken and the third was stuck in its location at unit 5 and 6. To complete a full cold shutdown of a nuclear reactor takes time and require power. It includes high pressure injection of coolant, depressurizing the reactor, injecting low pressure coolant, depressurizing and cooling of the containment vessel, and removing decay heat until the core is cool enough.

There are five barriers to prevent radioactive pollution of the environment from a Boiling Water nuclear Reactor (BWR). The first barrier is within the actual fuel pellets of compressed uranium dioxide powder, which in spite having more than 95% of the theoretical density of firm uranium dioxide contains void spaces where the fission products are contained. The second barrier is the fuel rod, which is an approximately 0.9 millimeter thick zirconium alloy tube holding a number of these pellets contained in helium gas. Fuel rods are combined 8-by-8 or 9-by-9 into fuel assemblies that in turn are combined into a 2-by-2 formation called a cell. Each cell has a cross-shaped control rod filled with boron carbide that absorbs neutrons thereby capable of slowing down and eventually halt the fission reaction. The reactor core will however remain very hot even if the fission is halted thus requiring continued cooling until the decay heat is below the boiling limit. In Fukushima the reactor core of unit 1 consists of 100 cells and the

cores in units 2, 3, and 4 consists of 137 cells. The third barrier is the reactor pressure vessel that extends to the secondary valve of each pipe connected to the vessel, and leakage of water or steam is not supposed to occur from this vessel. If the pressure becomes abnormally high, this must however be overlooked and pressure must be released to avoid a reactor pressure-vessel explosion, and for that purpose, there are safety-relief valves (SRV) that delivers steam to the pool in the pressure suppression chamber. The reactor core vessel is placed inside what is called the dry well and the containment suppression pool is the wet well. The two wells reside within the primary containment vessel that constitutes the fourth barrier. The fifth barrier is the reactor building, also known as the secondary containment vessel, which is designed to retain possible radioactive material that may have leaked out from the primary containment vessel. Both the fourth and the fifth barriers have similar solutions as the safety-relief valves of the third barrier because a so-called controlled release of radioactive material into the environment through valves and blowout panels is considered less severe than the uncontrolled release that would be the case if a barrier would explode.

If the water within the pressure vessel boils away, a Loss of Coolant Accident (LOCA) has occurred and the fuel will be damaged by the heat, which then becomes a core meltdown. Furthermore, the Zirconium alloy in the fuel assemblies will react in a self-accelerating manner with the water contained in the hot steam and release hydrogen gas. This gas becomes flammable at 4% and detonates at 10% atmospheric volume concentration. It is therefore essential to inject cooling water if need be, and for emergencies there is a dedicated Emergency Core Cooling System (ECCS), which however require electricity to function. To power the different parts of the ECCS system is one reason why the internal electrical power system of a nuclear power plant must have significant redundancy.

After the earthquake and the tsunami, to begin with it was considered that unit 2 was most critical, perhaps because its operational state was the most unknown. The Reactor Core Isolation Cooling system (RCIC) at unit 2 had been working after the tremors, but with the total loss of both AC and DC power systems that followed the tsunami it was suspected to have stopped. If the RCIC had stopped, the water level was known to reach the Top of the Active Fuel (TAF) by 21:40 the same evening, which would imply reactor core damage, the risk for this to happen was also reported to the government at 16:36. The prime minister issued at 21:23 a 3km evacuation zone.

However, unit 1 was actually in a more critical condition. The isolation condenser (IC), which is the main heat sink, had started to boil and the water level started to decrease. There was no backup water system due to the loss of power and the IC stopped working after it became isolated. Consequently, the cooling capability of unit 1 deteriorated rapidly. Unit 3 was on the other hand relatively well suited because the DC power had survived the tsunami, allowing the operators to read measurements, monitor the RCIC, and have

access to the High Pressure Coolant Injection system (HPCI). At around 22:00, a small portable generator made its way through the site and could provide basic power for lighting of the main control room for unit 3 and 4. For unit 1 the radiation level had around 21:50 increased and the building was declared off limits, it is estimated that the reactor core at that time had already started to be exposed above the water level and that the core had begun melting. The hydrogen build up must have been taking place already and the excessive pressure probably made gases leak out from the reactor. Presumably the containment vessel was beginning to exceed its design pressure and "Unit 1 was growing increasingly dangerous by the minute" (NAIIC-Ch2, p. 20).

There was no successful injection of water and at 02:30 on March 12th, the reactor pressure vessel of unit 1 failed and the pressure evened out within the containment vessel. Contaminated gas and hydrogen was constantly blowing from the pressurized containment vessel into the reactor building through cable penetrations, equipment hatches and such like. With the reduced pressure within the pressure containment vessel some water was injected successfully into the reactor using a fire-extinguishing pump, although with insufficient flow rate. Airborne radioactive material leaked out from the reactor building into the environment and the radiation level on the site increased. All workers were instructed to wear full-face respirators. This added further to the problematic work situation, for example within the main building and the main control room where the workers only had flashlights. The evacuation zone was expanded to 10km at 05:14.

The Ministry of Economy, Trade & Industry (METI), had ordered TEPCO to vent unit 1, but the valves were stuck and the working environment was becoming increasingly dangerous, thus, the progress was slow. At the same, time workers at unit 2 were desperately trying to route cables from generator trucks in order to get the Standby Liquid Control (SLC) system working, requiring extremely hard labor. When the cable routing was nearly complete, at 15:36, the hydrogen gas within the reactor building of unit 1 detonated and the building exploded. Five workers were injured, debris was thrown all over the place and the cabling was ruined. The shock was severe enough to knock off the blowout panels in the reactor building of unit 2. It was noted at a monitoring post at the site boundary that the radiation increased, making the Prime Minister extend the evacuation zone to 20km.

On March 13th at 02:42, the HPCI stopped at unit 3 and all means of water injection was lost. At 04:15, the core started to be uncovered and presumably then a massive amount of hydrogen developed. The operators dodged the extreme heat and went into the torus room to vent the reactor, a considerable challenge that actually succeeded. Batteries were then successfully connected to the safety-release (SR) valves allowing the pressure to lower enough to resume water injection, and the TAF became covered. However, the unit ran out of water at 12:20 and the TAF became uncovered again. At

unit 2, the RCIC continued to function but the operators started to prepare for a depressurization to allow for water injection by fire trucks because they estimated problems.

On March 14th, unit 3 boiled dry and at 04:30, the core had become completely uncovered. Fire trucks were preparing to assist with water injection when the building exploded at 11:01. Seven workers were injured, wreckage was thrown hundreds of meters high, and falling debris ripped a huge hole in the turbine-building roof. Seawater injection could not be resumed until after more than five hours. The explosion interrupted also once again the work at unit 2. Hoses and firetrucks were damaged and the workers had to start from scratch again. At 13:25, the RCIC at unit 2 stopped functioning and it was estimated that the reactor core would start to be uncovered by 16:30. Repeated aftershocks made, however, the work to be suspended until 16:00, and by 18:22, the core became fully uncovered. The SRV was released and it was discovered that the containment vessel probably was damaged because it did not keep the pressure. The fire trucks ran out of gas and the reactor continued to boil dry. More SR valves were opened facilitating more low-pressure injection of water that succeeded in keeping some water level in the reactor, but not in covering the TAF.

At 06:00 on March 15th, the reactor building of unit 4 exploded and a large noise was heard inside the torus room of unit 2, supposedly indicating a leak. Workers were removed and the monitoring of unit 2 stopped.

12.4.1 Accident investigation findings

It is not surprising that an earthquake of such magnitude as the Great Eastern Earthquake can damage a power plant, and perhaps even less surprising that a following tsunami can make things worse. Therefore, it is a trivial and uninteresting conclusion that it was the earthquake and the following tsunami that caused the Fukushima disaster. These great natural phenomena were the direct causes that triggered the disaster alright, but the interesting questions are whether these triggering events could have been foreseen, whether the disastrous consequences could have been prevented or mitigated, or if things actually could have turned out even worse. Nuclear reactors harness great powers, and with great powers follow great responsibilities. Governments and nuclear power producers of the world work therefore constantly with trying to take an appropriate responsibility and be prepared for whatever hazards that can threaten the safety for people in the world. Losing control of the processes within nuclear power plants implies to set lose powers that risk affect a great deal of the world, and “This accident, which has had a tremendous impact on Japan as well as on the world, is still ongoing” (NAIIC-Intro, p. 10). The difficult issue is what an appropriate responsibility means, and how inevitable uncertainties should be addressed.

The reports by NAIIC do comprise sections with in-depth analyses of specific actions and events for each of the six units at Fukushima respectively, but they state also repeatedly that no single person or operative role can be blamed for any specific consequence from actions taken. During these days and under these extreme conditions the stance of the Commission appears to be that everyone worked according to their expectations, and some much more than that. The Commission focused instead mostly on the responsibilities of the different authorities involved in the operation of the power plant. In particular, the Commission focused on the conditions under which the workers were performing their work, conditions principally determined by responsible operators and authorities, and that thereby could have been otherwise, if precautions for known hazards had been made or other design solutions selected. In fact:

> The Commission recognizes that the fundamental cause of the Fukushima nuclear accident originated from "the collapse of nuclear safety monitoring and supervising functions stemming from the reversal of the relationship between the regulators and regulated" among the successive regulatory authorities and TEPCO. Considering that there had been many opportunities for both sides to undertake safety measures beforehand, we regard that this accident was not a "natural disaster" but clearly "man-made". (NAIIC-Intro, p. 12).

The accident investigation commission states that the root causes of the Fukushima disaster "were the organizational and regulatory systems that supported faulty rationales for decisions and actions" (NAIIC-Exec, p. 16). The regulating authorities and TEPCO were since 2006 well aware of the risk for a total loss of electrical power if a tsunami would reach high enough, and the Nuclear and Industrial Safety Agency (NISA) was aware of that TEPCO had not taken any measures to remedy this risk. In spite of this knowledge, NISA had refrained from issuing specific instructions to meet these threats to public safety. In fact, the investigation commission found evidence that the relationship between the operators (TEPCO) and the regulators, in this case NISA and the Nuclear Safety Commission of Japan (NSC), was practically reversed (NAIIC-Exec, p. 16). The regulating authorities regularly asked explicitly for the operators' intentions when new regulations were to be implemented. As an explicit example, the report states that NSC informed the operators that the possibility for a station blackout (SBO, meaning a complete loss of electrical power) was negligible and therefore possible to disregard. NSC then asked the operators for a report providing the rationale why the risk for SBO could be considered negligible. SBO was, however, precisely was what happened on the 11th of March 2011. In addition, there was a negative attitude towards importing overseas advances in knowledge and technologies, which is one reason why the commission chose to label this disaster as "Made in Japan".

Then there is the issue of appropriate responsibility, which from both the perspective of the operators and the regulators appears to be considered limited to what can be foreseen. It seems the operator, TEPCO, and responsible authorities were quick to specify the tsunami as the cause of the accident, and downplay the impact of the earthquake. This was done by stating, "almost no important safety equipment was found to be damaged by the earthquake … to the extent that has been confirmed" (NAIIC-Intro, p.12), and, a "similar phrase has also appeared in an accident report submitted to IAEA [International Atomic Energy Agency] by the government" (NAIIC-Intro, p.12, brackets added). The Commission states as a plausible reason for these statements that:

> It may be construed, as an attempt to avoid responsibility by putting the blame on the "unexpected" (the tsunami), as in TEPCO's interim report. Through our investigation, however, we have verified that the people involved were aware of the risks from both earthquakes and tsunamis, and thus, there is no room for excuses (NAIIC-Intro, p. 12).

The NAIIC reports include also evaluations of operational problems, emergency response issues, evacuation issues, continuing public health and welfare issues, and follow up with suggestions for regulatory reformations, reformations of the operator, and reformation of laws, and so forth. These issues are however mostly beyond the scope of the present research, although one particular aspect appears worth mentioning, exemplified by the evaluation of operational problems. Initially questions were asked, could this disaster have been prevented, could it have turned out differently, and such like. Naturally this creates conclusions like, if this and that had been done, or if these instructions had been issued, or if certain technologies had been adopted, then things had probably turned out differently or certain stages of the disaster had perhaps even been prevented to happen. Such conclusions are the basis for lessons-learned knowledge and certain aspects of a necessarily perpetual safety work, but they are missing a crucial aspect. The aspect that the mere idea of considering sufficient safety achievable is framing safety measures within the domain of the predictable. This is the hallmark of calculative thinking where the generative perspective is missing, in this case by viewing safety from the outside, as a state. Safety must also be viewed from the inside as means of control that includes technology and psychology, in order for safety to continuously being achieved. Trouble is when the nature of desirable system workings, in this case a nuclear reaction, also is what is undesirable, a runaway nuclear reaction. It is a situation clearly comprising a tension between incompatible aspects.

12.4.2 Interpretations according to present frameworks

It has already been concluded in the analysis of the Forsmark incident that control of nuclear reactions is one of those things for which human beings require technological systems to manage. This has, among other things, the consequence that the domain of effects is strictly separated from the work domain (R1.1.1), implying that the character of controllability (R1.2.2), available feedback, and thus the situated understanding of the domain of effects is fundamentally dependent on systems design. The incentives for control (residing in the psychological domain, R1.1.5) are in practice determined by assumptions and conditions built into the technological systems. This situation in many aspects not at all different from other systems with technology mediated work domains such as airliners with abstracted system control, but there is a crucial difference in the character of coupling between domains. For nuclear power plants, the connection between local environments and the global physical and social world (R1.1.1) is significantly stronger than for airliners. While the crashing of airliners also may be triggered by unpredicted environmental conditions such as natural disasters or the mere presence of ice crystals and have long-term consequences on global traveling patterns and world economies, the direct linking to global matters is rather confined. Casualties in an airline crash may count to a maximum of a few hundred people and are essentially limited to those being on board the particular aircraft, perhaps including certain unlucky ones on the ground, if the crash site happens to be in an inhabited area. Although definitely a tragedy, the risk for additional casualties diminish quickly after the actual crash has taken place. On the other hand, when control of nuclear power plants is lost, lethal amounts of radioactive substances may spread over large areas and may remain for incomprehensible timeframes, and the number of casualties may accordingly depend on what happens in an indefinite future. This makes the nature of the actual situation during a nuclear disaster less important and shifts instead focus to the nature of overall matters. The focus for the previous cases was primarily on the character of controllability (R1.2.2) within the different situations, which is considered to depend heavily on assumptions about the nature of situations where control is supposed to be performed. For this case, it is more the assumptions about the nature of possible situations that require certain character of control that is the focus.

The Commission comes down with emphasis on the fact that TEPCO and NISA did not live up to international standards, and blames the Japanese culture of groupism for this disobedience of established foreign procedures. The international nuclear authorities had increased the norm for seismic resilience and the commission asks in the investigation report whether the disaster would have been less severe if these norms had been lived up to, a

valid question no doubt. However, the next level of concern is whether the new norms are sufficient or not.

To address directly the core issue, the considering whether certain established norms are sufficient or not indicates with unmistakable clarity a detached and calculative worldview that focus on predictability in systems assumed deterministic. In particular, the reasoning indicates an exaggerated belief in that calculative measures based on predictability are sufficient, which in fact is here considered the culprit, just as for the airliner cases. However, undesired consequences of exaggerated reliance on models are in this case working in the opposite direction concerning responsibility. Models and stereotypical assumptions are in aircraft designs built into automations and computerized abstraction layers thereby inhibiting pilots from taking their assumed responsibility. The inhibiting comes from system designs that effectively enforce the pilots to work in the realm of the predictable thereby making them unable to control the system comprehensively in unpredictable situations, which implies that they have no means for exercising their responsibility. For nuclear power plant operation models and stereotypical assumptions seem instead be used for escaping from assumed responsibility. Escaping is done by maintaining that unpredictable situations are impossible to know anything about, thereby not being possible to address. What cannot be known in advance is considered impossible to prepare for, and the responsibility for undesired consequences of such unpredictable events must therefore lay otherwise, an attitude by the NAIIC reports made perfectly clear as prevailing. This view implies to choose voluntarily to remain in the realm of the predictable and deliberately select what to take responsibility for. This stance is however to ignore completely the value of being prepared for the mere fact that unpredicted situations actually may occur, regardless how probable they are and what they might consist in.

No one can be prepared for everything, and one category of things that for instance insurance companies tend to refuse compensate for are so-called natural disasters, such as floods and avalanches, and, earthquakes and tsunamis. For nuclear power plants the possible consequences of such natural disasters are, however, severe enough to require some kind of preparedness, in order to justify their operation and ultimately make up the rationale for their existence. The problem is that predicting the magnitude of natural disasters and following courses of events is practically impossible, thereby making it impossible to know whether implemented safeguards are sufficiently strong. As for all man-made technologies with potentially dangerous properties, the problem always comes down to judging benefits against risks in terms of possibility and severity. When benefits are certain and dangers are manageable or sufficiently improbable, the judgment is easy. If the danger is extreme, on the other hand, even the smallest probability for it to happen makes the judgment of risks difficult.

It seems as that the severity of consequences associated with losing control of nuclear power plants is of such magnitude that this must never happen. To lose control over nuclear power plants is treated as inconceivable and practically impossible, yet a number of times it has happened, Harrisburg, Chernobyl, and Fukushima, to mention a few well-known cases.[84] The most important criteria for judging the entitlement for existence of nuclear reactors as justified then becomes that the possibility for anything unpredictable to happen that might lead to a loss of control is negligible. This is where the question of responsibility returns. It seems neither operators nor the regulating authorities are especially keen on taking the responsibility for the vast consequences of losing control of their nuclear power plants. The only way to justify their own existence becomes therefore to make sure that the possibility for this to happen is negligible and that the residual risk can be blamed on the inherently unpredictable thus being the responsibility of someone or something else.

Because of the above, it appears as in the nuclear power production business there is a high readiness for things to go wrong, but only for things that goes wrong in known ways and with known consequences. The risk for unknown ways and too extensive consequences is maintained as negligible and thereby not to be concerned about, in particular nothing to take responsibility for. Is not that exactly what a calculative safety culture is all about, a belief in the attainability of absolute safety, within certain well-defined limits of uncertainty? Perhaps is such a stance required for the ability to accept and be content with certain possible undesired consequences? There must simply be a negligible risk for unknown things to happen because the consequences are much too severe to cope with. The safety management model becomes thereby to prepare for known events until the unknown righteously appears possible to consider negligible.

There are several problems with this approach. First, the judgment of what amount of probability that righteously can be considered negligible is obviously arbitrary. Second, calculations of the probability for unknown events to occur are obviously impossible to do with any kind of validity because that would require knowing something about how much that is unknown, which is a contradiction in terms. The principal problem is, however, to consider anything negligible! It is the principal problem because to consider a threat negligible means actively to neglect preparing for it, which in turn means actively to make it certain that things will not work if the neglected happens anyway. The approach to nuclear safety in the present

[84]Harrisburg is, however, occasionally put forth as an incident with minor consequences, for instance in the Forsmark report (12.2), and as an example where the calculative safety measures with several barriers of protections succeeded in keeping the public from suffering. It is possible, though, to argue that this beneficial outcome instead is merely an exception to the rule that calculative safety measures actually might cause disasters (Reason 2000).

case appears as calculative as it ever can be and the result is so brittle it will snap at first sight of anything neglected.

When Fukushima first was built, the risk for earthquakes considered necessary to prepare for was judged to be earthquakes reaching a certain level on a specific scale, and the risk for a tsunami was negligible, perhaps essentially unknown. Safety measures to withstand earthquakes were implemented according to the judged levels on the selected scale, and measures to withstand a tsunami were consequently nothing but rudimentary, perhaps non-existing. The common knowledge about environmental conditions changed over the years and the judged risk for earthquakes to reach higher on the selected scale was increased, which made nuclear power plants in some parts of the world enhance their safety-measures to withstand earthquakes, but not in Japan, and not at the Fukushima power plant. As it were, the Great Eastern Earthquake occurred, which reached way above even the new internationally judged levels to withstand, and there was in addition a related tsunami. Perhaps is it true that if the Fukushima power plant had implemented measures to withstand earthquakes reaching the new levels on the selected scale, then the consequences would have been less severe. However, what had happened if the earthquake had been even worse and what about the tsunami? In particular, what had happened if another kind of natural disaster had occurred, something even less known about than a tsunami? For example, what would happen if the most fundamentally unknown kind of disaster had occurred, yet probably the most probable of them all, the deliberate act of sabotage by human beings with an unpredictable agenda? The point is, there will always be uncertainties, and it is plain stupid to neglect this fact.

The paradox is that in order to escape the calculative trap, it is necessary to stop disregarding the negligible. This might in fact imply to prepare less, at least to prepare less specifically. In order to make this position comprehensible, another metaphor might be required. Consider the preparation for a mountain hike on foot. If you would prepare for every situation that you know about that possibly could occur in the mountains, you would presumably require an entire truckload of equipment, effectively inhibiting the hike. What experienced mountaineering people usually do is that they prepare for the hike in general. What kind of challenges do you want to be prepared to meet, and what kind of challenges do you want to be able to spot and thereby avoid. The rationale for this approach is that if prepared for a specific challenge, chance is that it will not occur, but a different one will, thus render you unprepared. On the other hand, if prepared for relevant kinds of challenges, it is likely that if something resembling the prepared kind of challenge would happen, it is more likely that you would be able to dodge it somehow. Obviously this includes knowledge about own abilities, and especially about own skills in coping with certain challenges. In fact, the more skilled a person is in

mountaineering, the less equipment (i.e., technology) is usually brought along. This is also an example where practical wisdom (ch. 2.1.4) comes into play. At some point, most people realize that the planned hike implies too severe challenges making them opt for a less challenging route.

The moral sense of this metaphor is that it is necessary to stop aiming to prepare to a 100% for assessable hazards because this makes preparations unsurmountable, unlimitedly complex, and subject to economic balancing possibly leading to less harmonious preparations. In addition, it makes the prepared safety severely brittle. In reality, the calculative matter of either or is mostly irrelevant! The only way to be reasonably prepared for the unknown, is to be generally prepared (i.e., not perfectly prepared) for the known. By being generally prepared, there might be a chance that the general preparation also will be useful for similar unknowns. In fact, to be perfectly prepared for the presumably well-known will not only make you unprepared for the unknown, but probably also make you unprepared for the known because, presumably, the knowledge is imperfect.

In terms of the Fukushima disaster, this would translate to avoid being perfectly prepared for earthquakes up until a certain level on a specifically selected scale, but instead being generally prepared for earthquakes. It means to avoid neglecting consequences of a tsunami simply because the probability is considered negligible by someone. If the consequences of a tsunami are unacceptable but known to exist, preparations for handling a tsunami must be made, anything else is usually categorized mockingly as ostrich mentality (putting the head in the sand when things get tough).

For nuclear reactors the tension within the character of utility (R1.1.4) is painstakingly obvious. It is the very same property that facilitates the desired effects and risks creating far-reaching undesired effects, namely the extreme powers of nuclear reactions. On the one hand, these powers can be used to produce immense amounts of energy, and humankind seems always be short on energy. On the other hand, without sufficient control it may turns into an atomic bomb. While the fission-reaction set-up for most nuclear power plant reactors is considered unable to accelerate into a bomb kind of scenario, some other consequences of an uncontrolled fission reaction are equally severe as those of a rogue nuclear bomb, for instance the dispersion of radioactively polluted substances into the environment. The ultimate question of existence entitlement for nuclear power plants is left for others to consider (personally I am sincerely ambivalent), but the concluded causes of the Fukushima disaster makes it easy to doubt that humankind currently has the knowledge and powers to control these processes with sufficient safety and resilience among the inescapable uncertainties of the real world.

It all boils down to the problem of taking responsibility, and the tendency within the calculative culture to actively escape the taking of responsibility by hiding behind models considered objectively correct, thereby not the responsibility of anyone particular. In the Fukushima case, this tendency is

proved by the reaction of the operators and regulatory authorities to blame the tsunami, which was known to exist but categorized as unpredictable thereby a hazard not required to address. What has happened is that a belief in the ability to manage complex situations and only create beneficial effects by use of predicted idealistic scenarios and established regulatory systems, in reality turns out to be an exaggerated belief because situations are, evidently, more complex than predicted, which results in an inability to manage undesired effects. This regulatory lock-down to modeled assumptions resembles in fact the irony of automation (Bainbridge 1983). The irony with automation occurs when an exaggerated belief in predetermined and idealistic system workings results in technologies that become a problem despite having the purpose to be useful. Irony may be funny, ironic effects might be annoying, but the ironic effects of a regulatory lock-down implying a transfer of responsibility for undesired effects to the anonymous unknown is definitely a horror scenario.

13 Future technologies

13.1 Augmented reality

For the analysis of Augmented Reality (AR) systems, four scenarios were conceived, supposedly to illustrate four different kinds of use for AR technologies. These usage scenarios were: (AR1) in ground combat vehicles, (AR2) for dismounted soldiers in urban environments, (AR3) for maintenance of complex technological systems, and (AR4) for support of medical care. The scenarios were scrutinized by use of one SWOT[85] analysis-table each, followed by one summary conclusion for all four scenarios together. In order to comply with the structure used for the major cases in chapter 12, the layout will here be slightly different from the Fraunhofer and FHS reports.[86] First, all four scenarios are briefly recapitulated, then the findings from the analyses.

Combat ground vehicles (AR1) are commonly armored, implying that windows are weak points therefore avoided unless they absolutely are necessary. Compared to civilian vehicles this approach means that the visibility often is objectionably poor, and drivers are therefore sometimes forced to raise their heads through the manhole to get a better view for maneuvering, which obviously opens for assaults and other threats. The passengers have in addition no means to orient themselves while being inside the armored compartment and therefore they are easily confused when exiting the vehicle, which at the same time is one of the more dangerous moments in a hostile environment. With some kind of AR system for the driver, and perhaps for the passengers as well, the situation would be much different. The vehicle inhabitants would then not only be able to maintain their orientation under way, as while traveling by standard cars with normal windows. The outside view could in addition be complemented with specific task-oriented items of information from databases or command and control systems, or with information from sensors capable of seeing things human beings normally cannot see, such as infrared imagery. This kind of solution would doubtlessly provide a beneficial awareness about the present situation.

[85]SWOT-analysis is a method centering on a table with four fields labeled strengths and weaknesses (internal factors), as well as opportunities and threats (external factors).

[86]This chapter is based on the analysis made in the Fraunhofer institute report (Ruhlig 2011) and the Swedish National Defence College (SNDC) report (*Technology Forecast 2012 Military utility of ten technologies - a report from seminars at the SNDC Department of Military-Technology* 2012)

Soldiers in urban environments (AR2) face difficulties not present in a more traditional battlefield. Buildings and tunnels, networks of obscured roads and passages (covered from view by buildings), subways and sewer systems, they all have interior layouts that would provide an operational advantage if known about. An AR system could make this information visible in real-time. Just as for the combat vehicle, such a system in a networked environment could also convey task-related and other valuable items of information, for example data collected by non-local sensors.

Maintenance of complex technological systems (AR3), for instance military aircraft, requires a lot of documentation. Allegedly, the documentation and maintenance manual for the JAS 39 Gripen fighter aircraft is heavier than the airplane itself. By using an AR system the mechanic could have charts presented by the system, task-related components pointed out, and obscured component connections made visible. The dismounting- and remounting-procedures could for example be highlighted step by step, thereby presumably make them speedier while simultaneously guard against slips and damages that otherwise might follow from erroneous sequences of actions or from forgotten steps.

Support for medical care (AR4) resembles somewhat the maintenance scenario, but in this case, it is the maintenance of human beings. This medical scenario highlights one additional possibility of a networked AR system, the possibility to consult distant expertise. Specialists not present at the premise can take a more active part in the decision making if they have access to an AR system through which they can experience the medical situation.

13.1.1 Report findings

The purpose of an Augmented Reality (AR) system is to enrich the real world with virtual information generated in real time. It has four fundamental parts, or sub-systems; (AR-S1) a display capable of combining real world information with artificial information; (AR-S2) an enough accurate tracking system to allow virtual information to be properly aligned with the real world in the combining display; (AR-S3) a sufficient information source about the real world to be augmented, either locally managed or remotely accessed by a network connection; and (AR-S4) enough computational capacity to generate artificial information according to the tracked orientation of the display in relation to the real world.

The first kind of sub-system, the combining display (AR-S1), is perhaps the hallmark of an AR-system, and it is this special kind of displays that make them be information systems that augment the reality, and not only information systems about the reality. In principle, these kinds of displays could be developed for any one of the human senses, or for all of them. Although, displays combining artificial as in derived, represented, or

collected and redistributed, auditive and visual information with real world information are by far the most common kind of augmenting displays, and the latter (i.e., visual displays) the most common of these two, at least among systems specifically regarded as AR systems. That is, augmented visual reality systems are what one usually thinks of when considering AR systems. This conclusion is also confirmed by the fact that all four scenarios comprise augmented visual realities and the report maintains the view that visual AR-systems is the most common kind also in the future. The report distinguishes between two major kinds of visual AR displays, optical see-through displays that allow the user to see the reality directly and video see-through displays that present a more indirect view of the environment. Optical see-through displays have the advantage that they provide the user with a natural view of the real world while video see-through displays are better at handling occlusion effects between virtual and real objects. There is also a third kind of visual AR-displays, which are projection-based displays that present the artificial information directly onto the real world (e.g., a laser pointer).

Visual AR displays can in addition be divided in categories based on their design. They can be head mounted, then called Head Mounted Displays (HMDs), or hand-held, much like a modern mobile phone, or mounted in a vehicle, where the Head-Up Display (HUD) of fighter aircraft perhaps is the most common example, but today there are HUDs also for cars and maritime vessels. The hand-held solution has the disadvantage that it may become tiresome to hold the system and it limits the use of your hands, the head-mounted may likewise be heavy and obstruct natural vision.

The tracking system (AR-S2) can also be divided into categories: sensor-based, vision-based, and hybrid tracking. Sensors can for instance be electronic compasses, inertial sensors, or GPS (Global Positioning System) receivers, all feeding positional information to the AR system. Vision-based tracking relies on video images and calculation of position relative to captured images of real world objects, which in turn can be divided into feature-based tracking (e.g., databases of landmarks) and model-based tracking using (e.g., maps). Hybrid tracking combines the different tracking methods.

The requirement for computational capacity (AR-S4) is generally high for AR-systems because visual image processing continues to be a demanding task for computers. Together with a high requirement for dynamic real world synchronization, in combination with being lightweight and quickly moving, this becomes a true challenge. The report analysis conclusions are therefore based on a number of rather fundamental assumptions, such as availability of data in the form of digital maps and other harmonized spatial information, as well as sufficient data collections capabilities (i.e., sensors), communication and computation capabilities.

The potential benefits of AR systems are many and, in some sense, AR may be regarded as a reasonably mature technology. The combination of

navigation, blue force tracking (i.e., identification of and the knowing of own forces whereabouts), and target designation, simplifies and enhances operations. AR may facilitate group coordination of actions and risk avoidance. Local capture of intelligence required to feed the AR system may also be used by other systems as well, creating useful synergies.

Despite its maturity, AR technology still needs significant enhancements, for example with respect to simulator sickness that often comes from tracking, display, and computation, lags. Another problem, especially with carried AR systems, is that they add to an already heavy burden for dismounted soldiers, requiring further advancements in miniaturizing of technology. In fact, most of the problems with AR listed in the reports concern development challenges not yet fulfilled. However, the most severe problems are perhaps if the assumed conditions (e.g., availability of data, etc.) cannot be fulfilled and what happens if the systems fail when users have become used to them.

13.1.2 Interpretations according to present frameworks

The categorization of AR displays (AR-S1) can be used as a general note on the tendency to simplify things and reduce real-world aspects to be considered appropriately representable or even fully replaceable by technological properties. For example, video see-through are strictly speaking not systems augmenting the reality because there is no direct visual access to the reality as for optical see-through, provided that the field-of-view is sufficiently wide. Video images of the visual reality are merely reality conveying, these systems are not giving access to the real visual scenery. For video see-through displays, several natural aspects that we human beings make use of to perceive the visual reality, are lost. In particular, the direct connection between own movement (i.e., head and eye movement) and individually situated changes in the visual scenery is removed. Looking at a monitor (i.e., video see-through) implies looking at a two-dimensional picture regardless whether it is a computer generated image or a real-time view of the surroundings. This limit has direct implications for human understanding and awareness about the surroundings. Consequently, for video see-through kind of AR systems, the artificial overlays are then not augmenting reality but enhancing representations of visual realities, therefore better described as augmented artificial reality (AAR) systems. Personally I argue that this fundamental difference between natural eye-sight and an artificial conveying of a visual scenery prevails also for all kinds of 3D-imagery techniques known today, none of them are capable of mimicking all the cues human beings use for 3D-perception, and then especially not the connection to own body movement that is essential for situated understanding. As a side note, the kind of eye-tracking accuracy and computational power required for making human beings perceive artificial

images as a true visual reality are probably one of the greatest challenges as the human sensitivity for lag in this connection is extreme. My guess is that overcoming this lag is a much greater technological problem than reaching sufficient image quality. Consequently, my suggestion is that what matters is what kind of information that is presented on the display. The more lag there is in the artificial image, the more abstract should the information be. For example, if the target-designator symbol, which is related to the real world environment but still is an abstract kind of information, fluctuates a little in relation to your own body movements, it may be somewhat irritating. However, if the mountain depicted in front of you fluctuates in the same manner, you will lose confidence in the visual information and fail to act intuitively on the premise that there actually is a mountain there, a dangerous mistrust if you are a pilot.

The assumptions about availability of data and communication facilities indicate an exaggerated faith in stereotypical models and a calculative view of usefulness, despite that they in fact were acknowledged as fundamental. These assumptions make the entire analysis become calculative. For example, the urban soldier scenario (AR2) showed in the report an image of a battle situation where two soldiers ahead were marked with symbols for friends and a hidden enemy was pointed out by a target symbol, which clearly is a useful scenario. The usefulness is, however, obviously of the same character as the safety of a system considered handling all hazards, or following the same kind of reasoning as when stating that an accident would not have happened if a certain hazard had been known about, and so on. To analyze the usefulness of an information system and assuming availability of crucially valuable information is a give-away. A much more illuminating and probable scenario would be to analyze the usefulness of the AR system with the assumption that data is unavailable or erroneous.

The problem with artificially enhanced reality, regardless whether it is augmented visual reality or assisted vehicle control, is the inescapable mismatch between, on the one hand, the simplified and stereotypical models of reality on which the enhancements necessarily are based and, on the other hand, the real reality. In all cases where the augmentation addresses things that may differ locally from available system-internal maps of a generalized reality, the assisting system may, depending on its design, risk making things more difficult, just like the irony of automation (Bainbridge 1983). People must not only possess the skills to do things without the system if it, for some reason, fails to work as predicted, which then becomes more difficult after having get used to use the system, they must also know how to treat the system in order to make it fit the reality. The problem is all about 'the fog of war', it is about the fact that what in the end makes a difference is the ability to spot and handle deviations from the predicted, which requires, in the situation, the ability to get to know things not knowable in advance. For example, if a mechanic is using a maintenance AR system (AR3), how can

the system help in finding what is broken in the particular gadget worked on? Surely, the implemented map can only be about how the maintained system is supposed to be. A map can never show the way to a treasure hidden after it was drawn. The information revolution and the character of the information society has led to an increased focus on detached aspects (e.g., the maps) as the overwhelming flood of available information is mainly of a detached character. What really is going on, in a specific context or at a specific premise, continues to be the most valuable and hard to get information, perhaps especially within 'the fog of war'. When focusing on development issues, which always are about ensuring from a detached perspective that conditions become desirable effects, it is easy to forget the situated case of use within local conditions.

Another situated problem with the augmentation-model and reality mismatch that was omitted or disregarded in the report is what may be referred to as 'system lock-in'. A civilian example of the urban soldier scenario (AR2) was illustrated by a system solution that actually exists already today in various forms, video see-through augmented reality smart phone systems. By holding up your phone and view the world through its camera, in combination with having its position updated by its GPS-tracking facility, there are applications that can use the display to show an augmented image in which, for example, monuments, sights, street names, shops, and so forth, are marked. The problem named 'system lock-in' is that it frames focus to whatever is in the model and risk making the model more important than reality. While presumably having the virtuous goal of helping people experience more, an almost devious consequence is that people to a great extent probably will experience what is in the model, which presumably will be what those that pay for the model wants to be experienced. There are several calculative benefits clearly distinguishable from a detached perspective, while situated implications are much more subtle and difficult to assess. The most severe consequence from a military perspective is probably the risk for system lock-in to promote a stereotypical behavior because that is a very welcome effect from the perspective of an enemy.

Concerning the core component of AR technology, the display capable of augmenting reality, whether it is for visual, auditive or other sensory channels, there is presumably a lot of encouraging things to say (HUDs are great, HMDs too, provided that they are utilized cleverly). The present frameworks help to sort out what questions belong to what systems. To begin with, the overall worldview (R1.1.1), together with the outline and characterization of technological properties (R1.1.2 & R1.1.3), allow specifying what part of the AR system that is evaluated. The overall definition of usefulness (R2.2) connects system components with purposes. So, let us briefly examine only the AR display system (AR-S1). Presumably, AR-S1 systems are very good at enhancing situation awareness (R1.2.4) about entities that the information system actually knows about (AR-S3).

Hence, the AR idea seems in general productive to pursue, for situations and applications where availability of information is not an issue.

13.2 Nano air vehicles

The scenario conceived as a frame of reference when analyzing Nano Air Vehicles (NAVs) is a search and rescue mission within a building hit by mistake during a peace-enforcement operation.[87] Because the building is assessed as unstable, it is judged too dangerous for sending in personnel for reconnaissance, and a small swarm of NAVs is used initially. The NAVs are equipped with sensors such as stereo-video and infrared (IR) imagery together with some kind of communication capability. The video sensors are used for both navigation and finding of people, with IR as a complement and for spotting fires. Some NAVs are also equipped with microphones and speakers for the ability of rescue personnel to talk with trapped people.

When the swarm advances sufficiently far into the building the strength of communication signals reach their limits and one NAV lands at a suitable spot and turns itself into a relay station. Later on, the swarm divides itself into two groups entering deeper into the building in different directions, leaving a relay NAV whenever the signal strength becomes too low. The left-behind entities can also function as displacement sensors, sending out warnings for additional building collapses.

When trapped people are located, a NAV lands close to each one enabling the rescue personnel to communicate with them. Other NAVs have located hot spots and identified at least one as a potential fire. Throughout the entering of the building, some NAVs are trapped, they then inform the others and enter into relay mode.

13.2.1 Report Findings

By nano air vehicles, it is meant miniature aerial vehicles of sizes comparable to small birds or large insects. Due to their small sizes, they are highly sensitive for environmental conditions and thereby envisioned mainly to be used for military operations in urban terrain, within confined spaces, and indoors. This is in practice a completely new area for flying systems implying new challenges and opportunities. Their highly limited capacity for payload implies both limitations on the number of capability providing sub-systems and on the amount of energy available. The typical 'hover & stare' mission-profiles for larger systems are preferably replaced by 'perch & stare'

[87]This chapter is based on the analysis made in the Fraunhofer institute report (Huppertz 2011) and in (*Technology Forecast 2012 Military utility of ten technologies - a report from seminars at the SNDC Department of Military-Technology* 2012)

modes to save energy. The small size and low weight does however allow the system to stick to walls or land on small surfaces to perch.

The aerodynamic properties are rather different for small sized fliers compared to larger aircraft, a fact that both are facilitating and demanding. Natural fliers such as hummingbirds greatly outperform any artificially built flying system in terms of efficiency and speed relative own body length, undoubtedly desirable qualities. Development demands are in particular that it requires the development of artificial flying systems with flapping wings, which involves numerous technological challenges, including closed loop flight position control with suitable sensors, efficient flapping wing mechanisms, and enough capability of self-control (labeled autonomy).

For the operator interface, there are also certain challenges because in order to utilize fully the small footprint of the vehicle the control unit must also be small. The vision foresees smart-phones and head-mounted displays as control stations. The limited interface place additional demands on system self-management capabilities (denoted autonomy).

Presently there are two approaches for accomplishing NAV-systems, top-down attempts to miniaturize larger existing systems commonly utilizing rotating wings (i.e., helicopter style) or simple one-degree-freedom flapping wings, and bottom-up attempts that begin with novel micro-components and try to make them able to fly. The latter is by far the most complex approach and so far an area where there has been no success in realizing an airworthy vehicle with a complete set of necessary equipment (e.g., sensor, data storage and processing, energy, and radio communication).

The straightforward benefits of using NAVs are that they have the ability to access places otherwise inaccessible and, as for all unmanned vehicles, that they can reduce the risk for rescue personnel. They may facilitate increased situation awareness during rescue missions because of their capability to gain access to new areas with sensors and measures for communication. Problems are their limited range and sensitivity to environmental conditions as well as their dependency on automations.

13.2.2 Interpretations according to present frameworks

There is, at least in the military, a common saying that unmanned systems are most useful when things are dirty, dull, or dangerous, and these words may be interpreted rather widely. Dirty can be considered to denote all kinds of situations unpleasant or unsuitable for human beings, such as too small, too warm, poisonous, and the like. Dull may depict every kind of job in which human beings tend not to excel, deteriorate over time, or simply are not fit for the job, such as when there are repetitive tasks or when extreme or and repeatable precision is required, when the dynamics are too quick, for too long durations, and so on. Dangerous might denote all kinds of high-risk tasks, or even one-time missions, situations where the system is regarded

more or less expendable. In dirty, dull, and especially dangerous situations, systems not bringing people into harm's way are obviously useful, and for many dirty situations, there is in fact no alternative. The rescue-mission scenario above is such a situation where the technology adds an otherwise missing capability. To find technology useful that brings capabilities longed for is a trivial conclusion, an analysis of usefulness must therefore be sharpened to scrutinize more thoroughly how the capabilities affect the conduct of envisioned operations.

Without explicitly considering the situated future situation, and how the technology might influence the work, focus falls solely on the possibility of accomplishing the envisioned technology. Only by looking at how the technology is supposed to be used, it becomes possible to identify situated limitations, for example, issues similar to the system lock-in problems of AR-technology, and from such insights adjust the proposed designs in order to accomplish even better future systems. It is, presumably, necessary to look deliberately and with great zealousness for situated problems in order to design systems that rise above merely providing the missing capability and instead become a truly useful system.

The report do list explicitly the dependency on advanced automation as a potential threat against military usefulness of NAVs, but it appears as this threat mainly is considered a technology realization problem. The problem is framed to the issue that advanced automation requires high computational power or very good communication capabilities, which paradoxically for nano UAVs are aspects both counteracted by the limited size and energy supply of such small vehicles. The absence of implications for the situated perspective by use of stereotypical automations was noticeable.

13.3 The MODAS Project

Methods for designing autonomous systems (MODAS) is an innovation and research project involving a major Swedish manufacturer of long haul trucks as well as representatives from different Swedish universities (Krupenia *et al.* 2014). The project takes stance in knowledge about, in experiences of, as well as in empirical data from, current technology while addressing envisioned scenarios in terms of models of future driving situations and stating hypotheses about available technology. The future scenarios include complex driving environments with high and very high-density traffic situations requiring minimal vehicle separation down to sub-second distances as well as vehicle-to-vehicle and vehicle-to-infrastructure communication facilities.

In the early stages of the project, it was stated that from a driver's perspective, the ability to operate a manual vehicle under the envisioned conditions would be difficult or impossible, or even legislated against. As a

remedy, autonomous in-vehicle systems were thought to provide an opportunity for the driver to survive in the highly complex traffic environment of the future. Thus, initially, the expectancy was that autonomous in-vehicle systems or semi-autonomous sub-systems could provide the solution to the problem of an all too complex environment for future truck drivers.

13.3.1 Current investigations and continuation of the project

The MODAS project has a clear focus on human factors, including the insight that design concepts represent hypotheses about the relationship between new forms of automation and human cognition/collaboration, which are tentative hypotheses that need to be tested empirically and that must be open to revision as long as one learns from the mutual shaping that goes on between artifacts and human actors (Dekker and Woods 2002). Another input to the project is the GMOC-model (e.g., Jansson *et al.* 2006) of human decision making in dynamic systems, a model that can be seen as an applied version of the dynamic decision making approach proposed by Brehmer (e.g., 1992). After the project started, perhaps in part because of considering GMOC, the group realized the need to scrutinize some assumptions about the future and the meaning of certain concepts. Four issues were identified as necessary to address:

1. How does the concept autonomous relate to topics like automation? Is an autonomous system different from a fully automated system or not?
2. What philosophy of automation should the project build on? How do we make the best out of technology without losing human authority?
3. How can the GMOC-model be applied in the case of long-haul driving and how does it translate into a method for systems design?
4. What conceptual design solutions can be derived from the initial analyses and how should they be subjected to empirical testing?

Empirical investigations were conducted that included observations of truck driving today and interviews with truck drivers. The drivers were shown four types of automations and they were asked for their preferences in relation to the future traffic-situation scenario with very high density and minimal vehicle separation and a lot of communication going on between the vehicles and the infrastructure. The different types of automation shown to the drivers extended from supporting information systems and augmentation of the environment, support for recognition and interpretation, suggested action decisions, to actions implemented by the truck itself or its in-vehicle support systems. Design concepts, that is, design hypotheses, were developed according to conclusions drawn from observations and interviews. These design proposals were supposed to support the information needs of the

drivers and presenting the information in a way that enhance development of mental models such that they become sufficient for the driver to maintain control of the vehicle. Two design concepts were included in user tests focusing on observability. The aim of the study was to gain information of whether the concepts support driver understanding of different situations and enhance driver ability to regain control, for instance, in the event of automation failure (cf. AF447, ch. 12.3).

The result of the MODAS-project so far show that when empirical testing starts and real drivers are used as evaluators, design concepts change. It is not the question of performing a user-centered analysis, which tend to adopt current practices, it is neither a descriptive analysis merely concerning how users perform today, nor a normative analysis finding out how they should perform, it is more of a formative approach identifying how the interaction could work. An important conclusion is that, by the end of the day, the driver is the one that is responsible for all actions taken during driving, regardless the level of automation and implemented technological functionality. The level of automation is in the end an issue of authority and responsibility. Consequently, while the project acronym is retained, the use of the word autonomous for sub-systems and technological functionality has decreased.

13.3.2 Interpretations according to present frameworks

Initially, the project may be characterized as a typical technologist approach in which engineering solutions are assumed to solve most problems, sometimes this approach is implicitly assumed and sometimes explicitly stated. For a technology manufacturer the approach is natural and creativity facilitating because many employees are engineers used to think in terms of technological solutions to identified problems. However, it is an approach governed by a detached perspective focusing on technological feasibility and forgetting about situated aspects and responsibility. The project name contains also the phrase 'autonomous systems' and it appears as little reflection have been made on what actually was meant by using the phrase. Presumably, selection of the phrase was made because it was considered to denote an envisioned desirable increase in technological capabilities of future systems. The phrase appears actually to be a current buzzword among developers of high-tech systems (cf. ch. 15.1), for example, among developers of unmanned aerial vehicles (UAVs).

Later in the project, however, when user testing began and when human factors researchers joined the work, the involved perspective appears to have become considered increasingly relevant, indicated by the listed issues, and particularly by the concerns about authority. This development is, arguably, a tangible proof of how easy it is to adopt a detached perspective for systems designers not personally involved in using the systems to produce envisioned beneficial effects. Just as for AR technologies (ch. 13.1) and NAVs (ch.

13.2), usefulness was implicitly taken as a direct consequence of technological feasibility, an approach challenged by the involvement of human factors researchers. How the project will end is, however, impossible to know. Will the development continue to increase its focus on involved aspects or will concrete feasibility regain its status as the primary focus? In light of the development for airliners and nuclear power plants (ch. 12), a frightening but unfortunately not that far-fetched scenario is that legislating authorities, which inescapably are thoroughly detached, will continue to confuse predictability with safety and enforce implementation of highly automated vehicle systems that are lacking necessary means for situated control. The authorities are thereby removing authority from drivers and transferring it to computerized technology. Autonomous human thinking is then effectively regarded of lower value than automated logics.

14 Phase two conclusions

14.1 Addressing RQ2, the case of use

Evidently, in all the studied main cases (the future cases excluded) important performance edges were crossed with loss of control as a direct consequence, indicated by the fact that the events actually developed into severe incidents or tragic disasters. The interpretations according to present frameworks presented in chapter 12 are summarized (see Table 14.1 – Table 14.4 below) in terms of the two questions directly addressing RQ2 – The case of use (also presented in Stensson and Jansson 2014b):

- Did the crew/team see the performance edge coming?
- Did they have situated controllability, that is, did they have means to intervene?

Situation Awareness	**The Four Cases**			
	SK751	**Forsmark**	**AF447**	**Fukushima**
SA Level 1	Yes	Yes	Yes	Yes
SA Level 2	Yes	Yes	No	Yes
SA Level 3	Yes	No	No	No
SA – Summarized	They kept (with help from the assisting pilot) their awareness about the situation throughout the event	They noticed the loss of external power and understood its consequences, but could not anticipate – caused by an external mistake	They noticed that the situation was unfavorable, but they were, at least not until it was too late, aware of that they were stalling	They knew about tsunamis and about the possible impact on the power plant, but ignored the situation because the risk was judged insignificant

Table 14.1: Summary of the four cases, part 1: Situation Awareness

System Awareness	The Four Cases			
	SK751	**Forsmark**	**AF447**	**Fukushima**
SysA Level 1	Not really	Yes	Yes	Yes
SysA Level 2	No	No	No	Yes?
SysA Level 3	No	No	No	Perhaps?
SysA – Summarized	They did not understand the engine regulator sub-system and were therefore not aware of the slowly increasing thrust settings	They knew about the UPS units, but did not comprehend how they really worked and could therefore not anticipate the failure to connect	They knew about the flight-control automation systems, but they did not understand system limits and could not anticipate mode-shifts	System awareness is largely irrelevant for this case as it was overwhelming situational aspects that led to loss of control

Table 14.2: Summary of the four cases, part 2: System Awareness

Edge Awareness	The Four Cases			
	SK751	**Forsmark**	**AF447**	**Fukushima**
EA – Did they see the performance edge coming?	No, the engines started pumping and there was no (sufficient) control input to stop it	No, the UPS automation failed silently	No, not the stall edge, nor the flight-control mode-shifts	No, they worked under the assumption that it would not happen
EA – Did they have situated controllability?	Yes, the automation could be overridden by simply shifting the thrust levers by force	Not intentionally, but there were additional means of manual control that were utilized	Yes, in principle, the aircraft was in manual mode but the pilots were not trained to use it	No, required means for control under such conditions did not exist
EA – Summarized: **Action Regulation, based on SA+SysA+EA**	Out of the loop, they were unaware of the existence of the thrust regulation automation, thereby never really within the engine thrust control loop	Out of the loop initially, but later they regained control	Out of the loop, by being content with the idea that the aircraft could not stall the pilots willingly remained out of the loop – unable to gain control	Initially in the loop, until the situation escalated beyond control

Table 14.3: Summary of the four cases, part 3: Edge Awareness

Conclusions	The Four Cases			
	SK751	Forsmark	AF447	Fukushima
	Human contribution in airframe maneuvering (and luck) saved the situation. However, better EA for engine performance would probably have allowed for a normal landing with reduced engine power	Human contribution and sufficient controllability when finally realizing (from long time experience of working at the plant) the necessity to switch in the power generators manually saved the situation.	False safety depending on calculative models. An exaggerated belief in automated control implied inadequate practical means for situated control (inappropriate system design and training)	False safety depending on calculative models. The kind of situation that in fact occurred was considered ignorable

Table 14.4: Summary of the four cases, part 4: Conclusions

The overall interpretation after having studied these four real events and the three future scenarios is that the contemporary approach to controllability is calculative, with a calculative character of controllability as the result for designed technology. In the two studied domains (i.e., the air transport and the nuclear power production domains), the calculative control characteristic was a significant factor to what happened. The contributed frameworks and models are concluded to assist in realizing that this was the case.

For the airliners the calculative character of controllability is readily identifiable at a concrete hands-on system-control level, with implicit, or rather, quite explicit, notions of a calculative control culture at the organizational level as well. Because an airliner is a vehicle with a controlling crew, the fact that it ends up in a crash makes the crew obviously involved somehow and their control input (or lack thereof) significant to the outcome. Abstracted control characteristics according to calculative models and an adherent detachment from outcomes are therefore at the core. The exaggerated faith in detached and calculative control models had, particularly in the Air France case, spread also to those actually controlling the vehicle (i.e., the crew), as they seemed to be content with the idea that the computerized flight control system would prevent the aircraft from stalling. Supposedly, pilots consider themselves thoroughly involved in the situated control of the aircraft, even when controlling highly computerized aircraft, and they are in fact always involved, in the sense that they are physically located on-board the aircraft. However, the fact that the pilots apparently allowed themselves to be fairly content with their roles as detached high-level controllers is, arguably, indicative for a general shift in

concept meanings and an overall reduced richness for the view on the character of controllability. While focusing on physical control of a system according to well-known deterministic aspects (a task for which automation is well suited), the traditional role of pilots seems forgotten. Besides having the ability to handle the aircraft under normal circumstances, it was earlier expected of pilots to be expert system handlers in general. The backup task for pilots, or sea captains, bus drivers, etc., sometimes not outspoken, but always assumed, has traditionally been (arguably) to have the role of maintaining necessary awareness as well as sufficient knowledge and skills required for developing the incentive (e.g., by identifying anomalies) to diverge from the standard procedure, if an unpredicted unfavorable situation would occur, regardless the cause. In addition, it was expected that they then would have the ability to go through with such execution of non-standard control. It was at least assumed that there would be no one else better suited for the task. In short, it was expected of pilots to act as skillful autonomous system controllers taking responsibility for system behavior, beyond the domain of the predictable. A typical example when such taking of responsibility leads to a desirable outcome is "the miracle on the Hudson" (National Transportation Safety Board 2010) mentioned in chapter 12.1, where the captain succeeded with the heroic emergency landing much by having enough experience to know what formal procedures to skip. Today, on the other hand, the concept of controllability has essentially been reduced to the calculative level of computational models, and the expected behavior of human controllers is to act only according to predetermined procedures, a role that in fact is better executed by automation than by human beings. However, what happens if pilots actually become fully replaced by automations? What will happen when the stereotypical assumptions that have been implemented as control models clash with reality? When no one have the means to do anything about it, who will be held responsible for the outcome? The unforeseen situation will occur eventually, it is the normal constitution of complex systems. Without situated controllability, the couplings are as tight as they can be and the normal outcome will therefore be accidents (Perrow 1999)!

For the nuclear power production events, the calculative character of controllability is more salient also at the organizational level. The exaggerated faith in that organizational models are sufficient to control for safe operation of nuclear power plants was explicitly designated as the cause of the Fukushima disaster. For Fukushima it was, particularly, the exaggerated faith in Japanese organizations that was identifies as the culprit, which are organizations characterized as almost blindly obedient to certain hierarchies. In the present context, the Japanese culture may be described as showing an exaggerated faith in organizational models. Indubitably, a faith in organizational structures does affect the character of technological systems as well because organizations often govern the development of the

technological systems they use, which in this case become the systems that the power plant operators are faced to work with. At the hands-on system control level, because of their nature, it is required that the control of nuclear reaction processes is highly abstracted. However, the calculative character of controllability that comes from the exaggerated faith in deterministic models is what makes the system tightly coupled. The exaggerated faith in organizational models is what connects the complexity of social aspects to tightly coupled physical system. Consequently, the system is tightly coupled system and of high complexity, which is a system that according to Perrow (1999) is bound to fail. In terms of resilience engineering (Hollnagel *et al.* 2006), the system shows a brittle character of safety, a character that actually may explain the Perrowian disaster prediction. For the most brittle of all brittle systems, if anything happens that falls outside the predicted, which eventually will happen, the tight coupling ensures that the system will fail at once. Hence, the brittleness comes from considering all hazards possible to be taken care of as actually being taken care of. That is, the most dangerous aspect of the calculative character of controllability is the considering of certain kinds of control possibilities as unnecessary because the need for them is considered negligible. As long as uncertainties about the future exist, which, presumably, always will be the case, at least until the daemon of Laplace somehow is made real, situated controllability must be an end in itself. To refrain deliberately from designing for controllability is, literally, to paint oneself into a corner. For a generative approach actually to become applied, the faith in detached and stereotypical models, regardless whether they are organizational or physical, must be moderated and the situated perspective rejuvenated.

By application of the contributed frameworks and concept definitions, in conjunction these cases illustrate undesired effects of a calculative character of controllability within the situation, in the case of use. They show that reality may bring combinations of interrelated sequential aspects that turn predictions on end and thereby call for unforeseen kinds of system control, which in turn require experience and skill in comprehensive system control. Means for human control is therefore a categorical imperative. The challenge is to make these means as beneficial as possible. That is, both counteracting predictable mistakes and common misinterpretations but without enforcing stereotypical model-based values.

14.2 Addressing RQ3, the case of design

The overall conclusion after having studied cases from the past, from the present, and from the future, is that the calculative approach to system controllability is prevailing, perhaps even escalating as a consequence of facilitated means for calculative controllability from advances in automation

and computerization. The character of controllability was calculative in the past cases and it is calculative in the present cases. Moreover, the future scenarios are saliently lacking descriptions of situated aspects. In fact, the future scenarios tend to equate possibility of technological advances and feasibility of future designs with usefulness, which is to disregard how these potentially beneficial technologies are to be applied within future real-life situations. The future cases indicate an approach to controllability (the second tension within the character of control, R1.2.2) implying that the situated perspective for the case of use will continue to be disregarded. The mode of analysis remains that of technical rationality, predominantly using hard systems thinking in which problem solutions become the setting up of conditions that are believed to ensure a desired outcome, which is to assume that the world is deterministic (R3.2, an exaggerated faith in objectivity and predictability). Human agency, the impact of active and situated decisions for actually achieving the desired outcomes, for avoiding undesired outcomes, as well as for adjusting system effects to suit local conditions and situated aspects, are then analytically made obsolete.

The escalation of the calculative approach is perhaps most evident in the airliner cases, presumably because they are examples with concrete system handling that make issues evident. In the Gottröra case, it was essentially an abstracted (detached and automated) control of the engine subsystems that made them break down and cause the crash. For the Air France case, this kind of abstracted control had been extended to include also the fundamental maneuvering of the airframe. In both cases, the abstraction was made such that it implied a reduction of system and edge awareness to the extent that responsible system controllers (i.e., the pilots) lost control of the system of systems they were supposed to control. In the nuclear power production cases, the escalation is perhaps not as visible, while the preservation of the calculative view is evident, presumably because the nuclear business has no other option (cf. the discussions of the nuclear cases, ch. 12.2.2 & 12.4.2).

For the escalating development of systems with a calculative character of controllability, the vicious circle culture (R3) with its corresponding reduction of concept richness (R3.1) was identified as a plausible explanation. The fear of uncertainty in combination with a desire for predictability and comprehensibility takes the form of an exaggerated faith in objectivity and the downplaying of situated human agency. These desires are two modes that interact in a mutually supportive manner and drive the development towards further implementations of predictable and apparently impeccable automations (according to models deprived of rich meanings and situated aspects). When essential concepts such as usefulness are reduced to denote technological properties and calculative aspects, the idea of perfect objectivity becomes in fact attainable, as it is often within the language of science (i.e., mathematics). Judgment and choice is then reduced to instrumental reason (Weizenbaum 1976) and to a mere calculation of

objectively correct options, which implies a flight from experience (Reed 1996) and an escape from responsibility (R3.4) because insightful and responsible decisions has become obsolete. Moreover, human variability and unpredictability is considered problematic instead of a desired virtue and is therefore often taken as another reason to implement predictable automations and computerized abstractions. Operators are not considered responsible enough, arguably because they are judged much by their failure to align with modeled behaviors, and they are therefore not considered to merit means for situated control. The situated perspective seems then not only forgotten, it appears in fact being actively opposed, presumably because it wrecks the calculative predictability of parsimonious models. The situated perspective is counteracted on behalf of the more highly valued objectivity of model predictions, despite the fact that all models are wrong (i.e., they are incomplete and simplified representations of reality, but the map is not the reality). These cases show that we are actively striving towards a synthetic 'hyper reality' (Baudrillard and Glaser 1994). The future cases indicate that this development will continue, and together with the impact that technology has on how we interpret things it is likely that the speed of this development will increase, unless usefulness and other concepts regain their richer meanings. The question is how much automation is too much automation (Hancock 2014)?

14.3 General conclusions

All cases get richer descriptions by use of the ontological frameworks (R1.1), and situated aspects of the course of events for the studied cases become more readily understandable by use of the generative frameworks (R1.2). In fact, the frameworks for generative aspects provide enhanced, and sometimes even alternative, explanations of what happened during these studied events. The future cases indicate moreover that conditions and attitudes will be similar in the future, implying a prediction (i.e., forming a hypothesis) that these explanations will be relevant also for future events. The cases show that we tend to analyze things mostly according to concrete and calculative aspects (i.e., by use of hard systems thinking and detached values) and forget about social and psychological aspects that depend on context and situated aspects, and these cases show that this is even more common when analyzing and assessing the future.

– Phase three –

15 Research implications

This chapter addresses the second research purpose, to reflect on the implications that the prevailing view of technological usefulness has on human autonomy, a view governed by the paradigm of technical rationality. As such, this chapter initiates phase three, an analytical corroboration of the theoretical contribution and an assessment of explanatory powers, although the corroboration efforts are essentially deferred until chapter 16. The following discussion is, rather, a set of reflections on implications for current practice, if the conclusions made in this research can be accepted (to the least as interesting rhetoric statements). If these conclusions are interesting, or even more, if they are considered plausible, then it should be possible to talk about research implications for practice coming from this research, that is, implications, taking part in 'morphogenetis' (e.g., Archer 2010). For example, if the conclusion is plausible that there is an ongoing reduction of concept richness for concepts such as usefulness and that this development is counterproductive for human emancipation, this would imply that a change of attitude is required, which in turn would be a significant research implication, would it not?

15.1 The meaning of concepts

When problematizing and scrutinizing the concept of usefulness the lack of situated aspects was identified as a prevailing and accelerating issue, presumably because of the still firmly rooted paradigm of technical rationality established by the industrial era. The contemporary view of usefulness was found much too often be viewed (i.e., defined and evaluated) from a detached perspective lacking involved aspects relevant for understanding real-life values. It was in addition found that the detached perspective tend to be regarded as the objective ideal thus resulting in a vicious circle of reduced concept richness in which the lacked situated perspective is further removed from focus. This development was identified as prevailing for several concepts related to human phenomena and qualities where the involved perspective is essential. In an effort to address the issue of reduced concept richness, a conference paper has been published focusing on the possible contribution of too hard systems thinking to this development (Stensson 2010). Moreover, one journal article has been

published concerning the ultimate concept regarding human emancipation, the concept of autonomy (Stensson and Jansson 2014a).[88] Today this concept is in some contexts used for technological systems, which, arguably, is to reduce its meaning to describe merely technological properties. In addition, the principal reasoning used to describe driving forces, underlying confusions, and undesired consequences for the concept of autonomy can be used analogously for other human phenomena such as intelligence, consciousness, awareness, perception, cognition, and so forth.

15.1.1 Intelligence, autonomy, and other human concepts

Kant referred to autonomy as the ability of human beings to reason as free agents without the influence of authority or inclination. This statement is based on the Categorical Imperative, the basic central philosophical concept of Kant's deontological moral philosophy.[89] Kant himself called this "the principle of autonomy of the will, in contrast with every other which I accordingly reckon as heteronomy" (Kant 1785). This view of autonomy is that of someone who supposedly is autonomous. The view is about the rights and obligations that come from being an autonomous entity. Heteronomy, on the other hand, infers that thinking is constrained by previous knowledge and authorities, rules and procedures, or biases and heuristics (Laaksoharju 2010, 2014). As an elaboration of this categorical distinction, it is also possible to view autonomy from two perspectives.

The original use of autonomy is to denote human individuals, groups of human beings, or entire societies as autonomous when they are capable of and allowed to decide things without external control. When used for entities other than oneself, this becomes autonomy viewed from the outside. Ideally, an autonomous state has a government that is able to act without obeying another state's wishes and an autonomous individual can decide whether to follow or refuse orders. Human autonomy is closely coupled with the right to be free and maintain personal integrity, dignity, and liberty (e.g., Morris 1980). Ultimately, being autonomous implies to have the intellectual power and freedom of will to stage mutiny against an undesired controlling force.[90]

[88]This chapter, as well as chapters 10.3 and 15.2 are essentially recapitulations of those parts of the article in Ergonomics (Stensson and Jansson 2014a) not covered elsewhere. The main difference compared to the article is that the journal version is focusing solely on the concept of autonomy, while these chapters aim to widen the reasoning to cover more explicitly also other concepts for human qualities, such as intelligence, consciousness, awareness, etc.

[89]"Finally, there is an imperative which commands a certain conduct immediately, without having as its condition any other purpose to be attained by it. This imperative is categorical. It concerns not the matter of the action, or its intended result, but its form and the principle of which it is itself a result; and what is essentially good in it consists in the mental disposition, let the consequence be what it may. This imperative may be called that of morality" (Kant 1785)

[90]This is in accordance with the encyclopedia definition of *autonomy*: "1 : the quality or state of being self-governing; *especially* : the right of self-government, 2 : self-directing freedom

Seen from the outside then, an autonomous entity is something capable of being self-governing and thereby impossible to control completely. This meaning of autonomy is also in agreement with the notion of an autonomic nervous system, the part of our neurological system that we (normally) cannot control by will.

The alternative view is to identify with the autonomous entity and, like Kant, consider autonomy from the inside. What does it mean to be autonomous in a social world too complex to be analyzed exhaustively as a formal system, with emergent properties and conflicting values? It means it is impossible to be aware of and scrutinize, especially beforehand, every possible alternative for every single choice, which would be required in order to select the (in reality non-existing) objectively most valid option. There is simply no such thing as an objective and indisputable truth for anything but trivial issues, making it practically unattainable to be always right. Consequently, being autonomous is largely about having the right to be wrong, which makes non-obvious decisions reduce to pragmatics and approaches considered feasible according to ethical standpoints (e.g., Rachels 2009). Being autonomous is having the right to select subjectively what to take into account and what to disregard when available information is overwhelming or contradictory. Furthermore, being autonomous is having the right to guess and use presuppositions when uncertainty prevails. That is, seen from the inside, being autonomous is to have the power and freedom (the right) to be subjective and act according to one's own will.

However, these rights come with certain obligations, and most importantly, they are justified only by taking responsibility. By exercising the right to be subjective, you become accountable for your decisions. Thus, you have to consider the implications of your actions. Between human beings, an innate mutual understanding exists about the importance of such considerations, ultimately grounded in the shared appreciation of human life, which appears as empathy and building of trust. Any violation of these obligations results in some level of refusal of your autonomous rights, from not being considered trustworthy by people you rely on to being physically deprived of your freedom by norm-enforcing powers of society. In other words, the necessary balance between rights and accountability is one aspect that makes the two perspectives inseparable. To have the right to act autonomously you must shoulder the obligations of being an autonomous accountable entity, which technology is unable to handle.

and especially moral independence, 3 : a self-governing state" (http://www.merriam-webster.com/dictionary/autonomy)

15.1.2 The current use of human concepts for technology

It is easy to realize how the division of human reasoning into autonomy and heteronomy applies to human factors and cognitive ergonomics. Designers create artifacts and systems and thus the principle of heteronomy always applies. In the case of human beings, the heteronomy principle applies but only sometimes, which is when we base our judgments and decisions on principles or heuristics (Kahneman 2011). In their daily activities, we expect people to interact with technology with reasonable consciousness and cognitive workload. It is when we are faced with unexpected, non-typical, or thoroughly unpredictable situations that we need to regain focused control over the situation. Suddenly, human operators find themselves not being able to rely on output from the artifact. Information conveyed through interfaces has to be evaluated within a different contextual frame, perhaps completely different from anything system designers have anticipated. The expectations are for people to perform extraordinarily well in these situations. However, to do that, human beings need opportunities to exercise their autonomous thinking, a thinking that is qualitatively different from any kind of response an artifact can produce. In these situations, we expect people to be able to exercise their responsibility for which we hold them accountable.

Unfortunately, current practice is different. A development exists, in which the word autonomous is used for robots and automated systems. One example is Gunetti et al. (2011, p. 1, both quotes), writing about "highly autonomous aircraft that are capable of carrying out a pre-planned mission on their own" and stating, "In fact, the increase of autonomy is one of the most recognisable trends in the UAV industry". Another example is Huppertz (2011, p. 6, emphasis in original), who uses the term for Nano Air Vehicles (NAV) and proposes that NAV "have to possess a significantly **higher autonomy** than existing UAV". Further, Müller (2011, pp. 5, 10, 10, respectively) used autonomous in describing an Unmanned Underwater Vehicle (UUV) project that developed and built "autonomous navigating biomimetic robotic fish", stressing the fact that "autonomous behavior and artificial intelligence are major subjects in the field of robotics" and that future "UUV will benefit from this". Sullins (2002, p. 2) is concerned about the ethical and moral status of autonomous robots, arguing that "the ethical status of autonomous robots, both as ethical agents and objects of ethical consideration, be based on, but not identical to, the ethical status of their makers, operators, and those people and other machines that will interact with them". Finally, Waraich et al. (2013, pp. 26, 27, respectively, brackets added) discuss the fact that "[a]utonomous UASes [Unmanned Aircraft Systems] fly a complete mission from takeoff to landing" and that "the evolution of UAS GCSes [Ground Control Stations] from ground to autonomous control also represent a dramatic change during the last decade".

One might think that this trend is quite specific for extremely hight-tech systems and industries, like the ones mentioned above. However, studies published in the journal Ergonomics show the same pattern of confused use of the concept. For example, Sauer et al. (2011) discuss levels of automation and they use the term autonomous for the highest level of automation. Bekier et al. (2011, p. 349) defined seven cooperation options characterizing different levels of automation, where the highest is the level where "automation autonomously decided the solution". Parasuraman (2000) equalizes autonomy with automation, stating that automation can be designed to relatively high levels of autonomy, levels above the 5th out of 10 levels of automation of decision and action selection defined in Parasuraman et al. (2000). Some authors use the term autonomous for specific tasks such as for robots traveling (Chen and Terrence 2009) and for movements, actions, and cooperative behaviors (Squire and Parasuraman 2010), where the latter authors named the robots semi-autonomous. Other examples where the term semi-autonomous is used are Chen and Terrence (2009), and Stanton and Young (1998). On the other hand, one example in which the term autonomous could have been used, but explicitly was not, is Balfe et al. (2012). The authors discuss the development of design principles for automated systems in train traffic control from a cognitive ergonomics perspective. Endsley and Kaber (1999) discuss levels of automation without using the term autonomy, even though the highest level of automation involves actions by the computer only. Some authors use the term autonomy to explicate degrees of control for human operators such as drivers (e.g., Funke *et al.* 2007). Sauer et al. (2011) use the term autonomy similarly, in combination with authority to describe that operator control can vary according to situated needs. Stanton and Young (1998, 2005) used the term to describe loss of driver autonomy, thus referring explicitly to human autonomy. Clearly, the perspective for which autonomy is used differs.

Normally, researchers in ergonomics and related disciplines do consider contextual and situational factors important, in particular for human components. It may actually be that researchers in these areas do not consider people and artifacts as identical agents. However, the recent general trend seems to be different. Overall, there is a development in which human beings and artifacts are expected to act as agents in the same human-artifact ensemble as a joint cognitive system in its own right (Hollnagel and Woods 2005). Hoc (2008, p. 74) stated "as machines become more and more autonomous, the human operators can consider them as cooperative agents and the human-machine cooperation framework would probably be more and more relevant in the future". Further, Kavathatzopoulos (2010, p. 9, both quotes) cautiously deliberated on the future possibility of having robots and systems as autonomous ethical agents, while also stating, the "possession of an original purpose, a basic feeling or an emotion", "the existence of an emotional base that guides the decision process", probably is the criterion for

being considered a really autonomous system. Billings (1997) discusses fully autonomous automation and suggests using a management by consent approach, rather than a management by exception approach. The question remains, what do people actually mean when using autonomous for technology? One should also be cognizant of how the meaning of concepts can change over time, and this is what is turned to next.

15.1.3 The vicious circle of drifting concept meanings

Several undesirable consequences of using concepts for human qualities in order to describe technological properties can be identified, regardless of whether technology can be made to have properties comparable with their human counterparts. One possible position in this respect is to consider the use of human concepts categorically unwarranted for technology, simply because artificial systems are not human beings. This position, however, is not the main concern, nor is it whether it is possible to design truly autonomous and intelligent machines, although this question has a long and venerable history (e.g., Dreyfus 1972, 1992, Dreyfus and Dreyfus 1988). The main concern is both these issues! The present purpose is to point out the vicious circle that develops from these two aspects in conjunction. Together, they lead to undesired consequences, despite the presumably virtuous intentions underlying the contemporary use of human concepts for technology and behind efforts to develop human-like technologies.

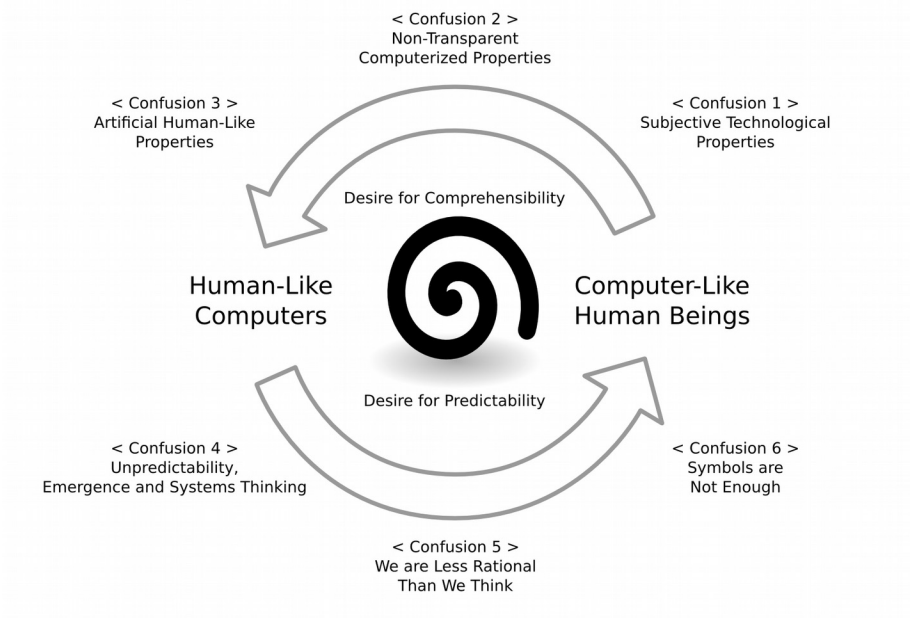


Figure 15.1: The vicious circle of reduced concept richness

Facilitated by the first three sources of confusion presented below and the natural human desire for comprehensibility combined with a tendency for human beings to understand things according to how we understand ourselves, we use brain metaphors for computers and credit technological systems with having human-like qualities such as intelligence and intentions. Facilitated by the remaining three sources of confusion and a strong human desire for clarity and predictability, we use computer metaphors for ourselves and dislike irrational emotions and highly variable human behaviors. We therefore try to force ourselves to act like computers. For example, when making low-level decisions, end-users are expected to mimic computers and comply perfectly with rules and procedures. End-users thus apply to the principle of heteronomy, sometimes in conflict with common sense or with overarching organizational goals such as efficiency or safety.

The computer-like ideal for human behavior and that computerized artificial properties are (implicitly) considered as being human-like are two modes that work together and mutually support a continuously impoverished meaning of concepts used exclusively for human qualities in the past (Figure 15.1). Thus, it is not just a question of the inappropriate use of the term autonomous for technological systems or whether automation runs the risk of reaching levels that principally cannot be externally controlled by human beings. Rather, it is about the interaction between these two developmental trends. As has been shown with examples and references (ch. 15.1.2), there are both scientific communities and commercial interests contributing to this development. The model in Figure 15.1 summarizes the six sources of confusion and the two modes driven by desires, as well as how the confusions interact in a vicious circle. The purpose now is to specify in detail how the six confusions manifest themselves.

15.1.4 Six sources of concept confusion

The first source of confusion (associated with the concepts discussed in ch. 5.3) is the tendency of users and developers to equal functional and machine properties with physical properties and treat them as objective. While in reality it is only the true physical properties, or rather, the physical aspect of functional and machine properties that might be viewed as practically objective. The implied technological subjectivity is however not a property of the machine, but of the designer and the user.

The second source of confusion (associated with the characteristics discussed in ch. 5.4) is that computerized technological systems have typical characteristics and properties that make intuitive human understanding of system workings difficult, leading to demands on system design perhaps not yet fully realized. Although the effects of poor system comprehensibility might resemble the effects from not understanding intentions of other people, this does not mean that the machine has a will of its own.

The third source of confusion (not explicitly associated with any particular model or framework, but instead a consequence of a lacking involved perspective, a central theme of the entire thesis) depends on the fact that the computer model of the brain appears alluringly easy to accept. Perhaps the special characteristics of computerized systems account for this favorable reception. We cannot observe the mind directly and therefore we sometimes become surprised about how other people think and behave (C1). Further, we cannot see what another person knows (C2), nor can we assess the abilities and skills of a person from appearance only (C3). The biological computer metaphor of the brain has been used with good explanatory power ever since the cognitive revolution, where the information-processing approach "provided a new way to study the mind" (Goldstein 2007, p. 13). For example, certain human characteristics become understandable if we think of our brain as having a limited number of input and output channels (e.g., Miller 1955), or as being equipped with a short-term (working) memory and a long-term memory (e.g., Goldstein 2007). However, one question is whether the strength of this metaphor comes from helping us understand the brain by relating it to a well-known technological system, or from helping us to understand the computer by relating it to the brain that we understand intuitively.

Although technological systems can have machine workings that make them behave similarly to living creatures, react consciously, and act intelligently (e.g., the robot dog Aibo),[91] this does not make them be conscious or intelligent. The Turing Test (Turing 1950), however, is often used to argue precisely that. One famous argument against this conclusion is John Searle's thought experiment, the Chinese Room (e.g., Searle 1980). The experiment shows that things can appear different from what they really are. Identifying something as intelligent even though it is known not to be intelligent would not prove the thing intelligent. Rather, it would simply prove the identification erroneous. Furthermore, perhaps Turing himself did not intend to show that computers could be intelligent. The interpretation may merely be that of his fellow computer scientists (Lanier 2011, chap. 2). Regardless of whether a technological system is intelligent or not, what we consider it to be determines our attitude.

It seems we human beings comprehend things according to how we understand ourselves, which is what creates premature expectations while interacting with human-like artifacts. If we believe a system to be 'intelligent' or 'conscious', we deceive ourselves by giving too much credit to machine properties that appear intelligent. We then intuitively infer the machine to have additional human-like qualities normally connected to our kind of intelligence, empathy for instance. The rather common reaction that a

[91]For instance http://en.wikipedia.org/wiki/AIBO

computer deliberately is causing you trouble when it simply keeps running its software exemplifies such intuitive inference.

The third source of confusion refers to computerized systems mistakenly credited for having human properties by use of a brain metaphor, or conversely, that the human brain is mistakenly described in terms of information-processing units by use of a computer metaphor. Metaphors feed on superficial similarities that indubitably also facilitate premature expectations for more similarities, which affect the relation between our view of technology and ourselves. Ultimately, the question is whether we want to consider technological systems similar to ourselves, or if we can benefit from explicitly discriminating between human beings and artifacts. The choice will determine expectations and to what extent we will trust and feel confident about autonomously acting systems.

The fourth source of confusion is essentially associated with chapter 7 and 8. Some fundamental assumptions of science are that a true physical universe does exist and that it is, in essence, an orderly system (Graziano and Raulin 2007, p. 31). Such assumptions make systems thinking an extremely powerful approach, mainly because of its clarity and calculability, which, in turn, may give an impression of objectivity following from the applicability of impeccable mathematics. For every specific problem, however, the universe must be reduced to a relevant system of concern in which all elements are identified with all their properties and interactions appropriately described. That reduction is a significant challenge in itself. By focusing too much on achieving a correct system description while having an incomplete set of factors because of not identifying a completely relevant system of concern, or failing to comprehend model shortcomings, a clear understanding may be achieved, but of the wrong problem. This circumstance is the ubiquitous scientific dilemma of the specific vs. the general (e.g., Boulding 1956) or rigor vs. relevance (e.g., Benbasat and Zmud 1999, Davenport and Markus 1999). It seems the rigor required for calculability counteracts the meaning required for relevance, and perhaps these aspects are fundamentally incompatible. The role of the concept of emergence in philosophy of science is to explain things that are “neither predictable from, deducible from, nor reducible to the parts alone” (Goldstein 1999, p. 57). Perhaps meaning is emergent in the system as a whole, available only at a level of abstraction above that of the calculable model? This idea is actually quite logical because the reduction, the identification of a relevant system of concern, is a subjective point of view. The system model is derived from what is considered a meaningful description, and consequently, the meaning of model findings is irrelevant without the considerations that fostered the model. This contextual dependency means that for anything but a closed problem of simplicity (Weaver 1948) with uncontroversial consequences there will be a significant amount of subjectivity involved, which inevitably implies ethical dilemmas

lacking objective answers. Thus, both models and model findings depend on moral standpoints.

The scientific ideal of objectivity can in this context be accused of reducing decision-making to calculative optimization according to current knowledge that tends to be a model of wrong problem. In fact, all models are wrong (Box 1976, Sterman 2002), making the value of gained knowledge fundamentally dependent on a humility toward its subjective grounding. Most problems occur because social dilemmas are wicked problems, according to Churchman (1967b), which makes the notion of optimal solutions irrelevant because there are no definitive and objective answers (Rittel and Webber 1973). Yet, the powers of systems thinking in combination with human psychology tend to frame interpretations of problems and decisions (Tversky and Kahneman 1981) to the mainly physical perspective of most models. However, what is actually the right problem is determined by what things mean to involved human beings and what they believe to be the right problem. This discrepancy between technological and social aspects is also reflected in the difference between hard and soft systems thinking in which a typical hard systems thinker considers complete systems as viewed by a human observer, whereas a soft systems thinker considers purposeful human activity systems (Checkland and Scholes 1999) that include the relevance perspective.

The fourth source of confusion is an effect of the analytical and rational ideal of science, as well as the adherent tendency to mistake the map for the territory. This tendency is labeled Platonicity, in which the understanding of things is framed by the seductive beauty of parsimonious platonic models, causing low-probability events to be regarded as inconceivable as the appearance of a black swan before it was known to exist (Taleb 2010). It is a matter of taking idealistic and too hard system models (i.e., closed and focusing on physics and calculability) as objective descriptions of the problem and giving them an exaggerated value by considering them as more complete than they really are. It may also be considered a systemic paradox (Stensson 2010) in which the more rational analytical effort required for the development of sufficiently complete models paradoxically leads to an obscured understanding of relevant aspects for the system as a whole. One consequence from actually taking the map for the territory by considering model implications more valid than real world effects is reduced concept richness, which, for example, is what analogously as for autonomy appears to happen with the concept of safety. Predetermined automatic behavior can be safe only as long as the models are correct, which we know they are never. Yet, human control is continuously being replaced with automation as a safety measure. This is a safety-paradox (Reason 2000) in which predictability is confused with safety, a situation that leads to stereotypical

safety measures that inevitably create disasters.[92] Safety must be a richer concept implying more than predictability. Analogously, being autonomous is more than being able to make decisions by consulting a map or follow a decision rule, which essentially is to be heteronomous. When technological autonomy is confused with human autonomy in this way, one might even call this an autonomy paradox.

The fifth source of confusion is the idealization of our rational and analytical capacity and the exaggerated reliance we put on it, especially the expectations about when and how we are supposed to use this ability. The problem is therefore not so much that we are less rational than we think, but rather our failure to realize this and design technological systems that we can interact with in a manner suitable for intuitive human beings.

The sixth source of confusion therefore concerns the case that, when people solve problems in unexpected situations, we often assume they do so in terms of processing symbols. In reality, however, decisions made by human beings are based on situated meanings and contextually dependent values. Suchman (1987), Klein et al. (1993) and Hutchins (1995) demonstrated the importance of cognition as part of action while Jansson et al. (2006), and Erlandsson and Jansson (2007, 2013) have shown how operators have situation- and context-dependent strategies in common, something that is possible only if they have been close colleagues and experts on the same control task for a long time. The fundamental difference between human beings and technological artifacts is the ability to evaluate symbols differently depending on the situation and context.

15.1.5 Consequences of concept confusion

The model illustrated by Figure 15.1 predicts two consequences, of which the first mainly concerns the meaning of concepts. Automation and autonomy may be considered overlapping concepts about properties from the same dimension. They both describe a self-contained nature, where automatic indicates a stereotypic mechanical behavior, whereas autonomous denotes self-governing in a more self-sufficient sense.[93] With this one-dimensional view, autonomy merely implies a more capable kind (a higher level) of automation, one that facilitates desirable qualities for technological

[92]For example, the Air France 447 crash (BEA, Bureau d'Enquètes et d'Analyses pour la sécurité de l'aviation civile 2012), that should be a wake-up call regarding automation and simulator training (Thomas 2011).

[93]This corresponds rather well with the encyclopedia definition of *automatic*: "1 a : largely or wholly involuntary, *especially* : REFLEX <*automatic* blinking of the eyelids>, b : acting or done spontaneous or unconsciously, c : done or produced as if by machine : MECHANICAL < the answers were *automatic*>, 2 : having a self-acting or self-regulating mechanism <an *automatic* transmission>, 3 *of a firearm* : firing repeatedly until the trigger is released" (http://www.merriam-webster.com/dictionary/automatic), while *autonomous* (in footnote 90, p. 244) is focusing more on rights and moral independence.

systems, which presumably is the rationale for the concept usage exemplified in chapter 15.1.2. This use of autonomy might therefore be dismissed as an innocent play on words, but I would like to stress that concept misuse might have severe consequences. Technological designs have a strong 'structuralizing' (Orlikowski 1992, drawing on, Giddens 1984) or 'morphogenetic' (e.g., Archer 2010) effect on activities because our understanding of things depends on our experience from activities (Reed 1996). Our experiences are at the same time expressed and shared by the concepts we use, and these experiences and concepts are what guides future designs. The meaning of concepts will affect system workings and structuralize our activities, and inevitably influence our understanding of things. This circular dependency, one of many in social systems, is what makes the conceptual framework used among designers and what things mean to them essential. Scientific paradigms and the meaning of concepts are tightly coupled (Hjørland 2009). Grounding assumptions and conceptualizations are particularly influential on (future) human activities during design of automated systems, primarily because such systems are probably the most structuralizing kind of technology that exists. The reason for the significant impact of automated technology on human activities is that these systems interfere directly and actively with what happens. They do this by applying their designed behavior in real situations. Because artifacts are heteronomous by design, the activity domain becomes increasingly calculative (Figure 6.1, p. 117).

From the perspective of human emancipation, a maladaptive use of autonomous for technological systems implies a counterproductive reduction of concept richness, manifesting as an impoverished meaning of a concept originally describing a rich quality exclusive for living things. Unfortunately, this development can also be observed for similar concepts for human qualities such as intelligence, awareness, and consciousness, a development that follows from allowing the aforementioned kind of word play to continue unbridled. This is problem one:

(P1) Reduced concept richness, when concepts for human qualities *de facto* becomes applicable to technological properties

Following from the vicious circle modeled in Figure 15.1 and from P1, a second problem arises. With a mix-up of human qualities and artificial properties and when concepts that could have been useful for explicating differences become applicable to technology from which explication is required, we as human beings are left without the means to maintain our position in relation to the artifacts. We thus find ourselves competing with technology on the grounds of technology and by rules stated by technology. Arguably, this is an unwarranted competition with unfair conditions. It is a competition on the grounds of technology because we very much like to

consider our ability to reason logically like computers our hallmark quality, but when it comes to formal processing of symbols we are since long outperformed. The rules are set by technology because we have idealized the grounds and consider the design-governing artificial worldview the true reality (i.e., we prefer a clear-cut map before a disorderly reality), despite that this ignores the richer perspective determining the true values. It is an unwarranted competition because the artificial territory is irrelevant without knowledge about the true values fostering the model, something artificial things cannot understand. However, with idealized rules within idealized grounds, an objective processing of symbols becomes impeccable. The natural result, if calculative rationality is allowed to govern judgment, is that we identify our emotionally grounded values inferior to the calculative logic of computers. We then surrender unconditionally, choosing to fall behind and let the computers decide, which is to surrender our own autonomy. This is problem two:

(P2) Reduced human autonomy that arises when technology *de facto* is granted increased 'autonomy' (i.e., decision authority)

These two problems (P1 and P2) relate to a development that is taking place today, apparently without sufficient scientific discussion. The problems constitute a major challenge for research in a number of areas, such as human-computer interaction, cognitive ergonomics, human factors, operations research, management science, and military operations. The purpose of this passage is not to discuss all the implications that follow from the model. Instead, implications for the human role are summarized.

15.2 Human authority, a categorical imperative

In the end, human beings are responsible for their actions to the degree that they can be brought into court for violating certain norms or procedures, or they can suffer personally from the consequences of their decisions. People risk losing their societal rights, their freedom, or ultimately, their lives, by making wrong or poor decisions. Technological artifacts, on the other hand, do not run that risk. One consequence of these risks is that most (if not all) decisions human beings make are based on situated meanings and contextually dependent values ultimately grounded in an innate knowledge about the value of human life, which is constantly at stake. Technological artifacts do not have a life of their own and thus can never know the real meaning of fundamental human values. Accordingly, decisions made by artifacts will be stereotypical, slightly off-target, or simply miss the point, making human authority over technological workings a categorical moral imperative. The true meaning of things, real values, and relevant purposes

are all emergent features of the context, determined in the actual situation by involved people (e.g., Dourish 2004). This fact makes decisions by people truly involved more relevant than artificial decisions, regardless of whether they are model-compliant and thereby considered objectively correct. That is our role, a role we cannot escape. Our role is to make sure activities are relevant and justified because while the artifact may seize to exist, we are held accountable for its effects, either while alive or posthumously. For this role, people require to be sufficiently involved emotionally to care and have the authority to adjust activities according to contextually situated interpretations of actual situations.

The problem is that maintaining such authority gets more and more difficult as system designers seek technological autonomy, regardless of whether the aim is metaphorical or genuine. The reduced authority occurs in any case because even a metaphorical aim has the intent to remove (the need for) human intervention (why else the metaphor?). However, technological structuralization transforms this aim into a self-fulfilling prophecy. To maintain authority, and if necessary enforce the control of activities, there must be proper means to intervene. For human beings, means for intervention must be accompanied by means to develop the incentive to intervene. Moreover, to develop the incentive requires sufficient knowledge about system workings to identify why and know when to intervene, as well as sufficient skills to be confident about how to intervene. Incentives, will, purposes, interpretations, meanings, emotions, values, etc., are all aspects of the psychological domain, aspects that are contextually situated and practically inseparable from the activity domain. This inseparability occurs because we learn to know things and develop skills through activities (again Figure 6.1, p. 117). However, the activity domain is becoming annexed by autonomously acting technological systems explicitly designed to be self-sufficient and independent of human guidance.[94] This alienation from the activity domain obstructs the development of skills and the attainment of knowledge because to understand what goes on, we have to be actively involved. There are several authors discussing the issue of automation intervening with the contextual domains of activity and psychology (e.g., Stanton and Young 1998, 2005, Endsley and Kaber 1999, Parasuraman 2000, Young and Stanton 2007a, 2007b, Röttger *et al.* 2009, Squire and Parasuraman 2010, Bekier *et al.* 2011, Sauer *et al.* 2012, Jansson *et al.* 2014). The deliberately accomplished non-involvement does not only remove the need for intervention but also inhibits the development of incentives for intervention and therefore the prophecy that human control is obsolete will be fulfilled. Teleoperation and the contextual distance between design and use of automated systems are other examples where indirect or

[94]Illustrated in Figure 6.1 p. 117 by automation raising the boundary between the calculative and the contextual.

detached human involvement creates a tendency to align behavior with idealistic models. The alarming aspect of this development is that our contemporary predisposition to prefer the map before reality causes us to conclude that the resulting absence of involved human intervention is proof of the validity of the artificial behavior. In practice, what happens is that along with the reduction of richness for concepts such as intelligence, awareness, and autonomy (P1), activities are being deprived of their psychological connections and decisions regarding real-life values are reduced to the symbolic level of computers (P2). The psychological domain then becomes obsolete. To this, I object with emphasis!

Our role is, individually or as a contextually relevant group, to decide autonomously what should be done, which commonly include the use of technology. The role of technology is to do what we want it to do, the way we want to have it done, which is to be heteronomous. The reduction of autonomy to mean merely the ability to make autonomous decisions (within certain limits) is quite different from the intuitive interpretation most people make. In general, people think about autonomous beings, simply because we tend to understand things according to how we understand ourselves. The difference is crucial and a change of concept meaning would be paradigmatic and counterproductive because it disregards the situated psychological domain of activities. Arguably, how we intuitively interpret concepts is what really matters. We seem to act initially on our intuitive interpretations (Kahneman 2011), and intuitively, we assume autonomous beings to incorporate the psychological domain, show empathy, know about obligations, and take responsibility for their actions, all because we trust people to know what it entails to have their lives at risk.

The fourth source of confusion, the systemic paradox, constitutes a keystone for the issues discussed here. When the map is taken for the territory and real-life activities are enforced to comply with behaviors a priori judged as safe and effective, the result is stereotypical and context ignorant activities that risk being irrelevant and hazardous. There is an ongoing "perverse and fruitless pursuit of technological solutions", a prevailing "science fiction fantasy that technologists can somehow automate all critical human functions" (Lintern 2005, p. 402, both quotes), which is justified because consequences become invisible if the world is reduced to a platonic map. The powers of systems thinking are important for understanding how things are likely to work if triggered, but they say little about why these workings should be triggered or when it is desired and appropriate, simply because traditional (hard) systems thinking is not very well suited to describe such emergent aspects of social systems.

The fifth source of confusion is catalytic to the fourth. The ideal of objective rationality and the worshiping of our own ability to reason logically becomes in conjunction a double-edged sword when we become cognizant that we are less rational than we think. When we identify

rationality as our hallmark quality and find ourselves inferior to technological systems regarding formal logics, the reaction to resign and let the computers take over is somewhat understandable. What is disregarded is that objective rationality is a devious illusion. First, it is practically unattainable, and second, it is detestable in that it implies a stereotypical and meaningless order.

The sixth source of confusion is grounded in the fourth and fifth. Symbols are the crown jewel of computerized logic and formalized rational reasoning. When used, they quickly become one of the most powerful tools for concealing the true meaning of things under the guise of objectivity. The danger lies in the very nature of symbols. For a symbol to be useful, it must mean something, but it must also be valid in general to some extent. Otherwise, there would have to be an infinite number of symbols, causing each one to become less significant. Symbols are therefore never enough because they separate defining aspects from their contexts, which, paradoxically, deprives them of their true meaning. That is why computers, which are exclusively symbol interpreters, consequently and continuously miss the point when making decisions. To provide true meaning symbols are required to be interpreted within a real situation. The problem is that for human values this is something only human beings can do.

These three sources of confusion (i.e., the fourth, fifth and sixth confusions) are in some sense only one source. All three are about mistaking a largely unpredictable world of subjective meanings for an objective and predictable machine, and about considering that view the ideal. This combined confusion is driven by a completely natural human desire, namely the desire for predictability. Unfortunately, the desire is strong enough to make us believe too much in our calculable system models, perhaps because desired predictability otherwise would be hampered. The first three sources of confusion are in some sense also only one source, but about the opposite. They are about mistaking calculative machines for subjective and informed beings. This combined confusion is also driven by a completely natural human desire, namely the desire for comprehensibility. Unfortunately, this desire is strong enough to make us infer properties on the machines that machines do not possess, perhaps because these properties belong to our intuitive understanding. Together, these desires and confusions form a vicious circle of continuously interacting aspects that contribute to each other's momentum, as illustrated by the model presented in chapter 15.1.3.

However, it is paradoxical that machines are credited for having human-like properties when we discredit ourselves for having such properties. When technological workings are unpredictable because functional properties are more subjective than we think (confusion 1) and system workings are incomprehensible because of non-transparent model-based behavior (confusion 2), we consider these workings an asset that make artifacts intelligent and human-like (confusion 3) and possibly autonomous. For our

own part, we regard such subjective unpredictability a weakness and a reason to let artificially intelligent technological systems control our activities, a decision that necessarily impedes our own autonomy! How can superficial human-like properties, which are poor and stereotypical representations of the real thing, be regarded more valuable than real human properties? Logically, this can only be true if concepts originally describing human qualities have been deprived of their richness and now referring only to technological properties. Obviously, technological systems provide better technological properties than human beings do and hence the real culprit is the problem of reduced richness of concepts for human qualities.

Roughly, there can be three reasons for why people use the term autonomous for robots, systems, and technology in general. First, a person may be unaware of the actual meaning of the term, just using it because other people use it under similar circumstances. Second, people may be using the term in a restricted sense, focusing on a limited part of its meaning. Third, the term may be used because it is believed to describe something desirable, something not yet accomplished with technology in its current state of development. The first reason is unfortunate, perhaps especially for the person using the term who may not fully understand the implications of the conveyed message. The second reason is problematic because the simplified usage leads to a reduction of concept richness (P1). The third reason is a usage that comes close to the Kantian understanding of the term in which it becomes counterproductive because if accomplished the supposedly desirable effects of autonomous technology implies instead a relative impeding of human authority (P2).

15.3 Philosophical reflections

How is it that the present seems so often forgotten or disregarded? For concepts such as usefulness, effectiveness, and safety, the predicted theoretical future seems often regarded more important to assess than the real world now, despite the fact that the future depends on what happens now. Why is it so? It may be possible to find an answer to this question in how we currently tend to view the world, scientifically. Perhaps is the root cause of the lost situated perspective a philosophical issue? The following discussion is an effort to continue the philosophical reasoning initiated in chapter 2.1, in light of the insights gained from this research. The situated presence is lost presumably because of contemporary philosophical assumptions reinforcing technical rationality. A philosophical discussion is therefore required for the possibility to transcend paradigmatic boundaries. The problem with the lost perspective concerns, arguably, not so much potential misinterpretations of the world, but more how these interpretations affect the view of ourselves and of our role in the world.

15.3.1 Scientific abolition of the generative present

In essence, the ideas presented here are based on that the situated present perspective is lost in scientific analyses mainly as a consequence of what here is designated **the systemic gap** (Figure 15.2 – A, the vertical gap between the past and the future). It is a gap formed by systemic decomposition and formal descriptions of reality, often as a consequence of using the language of science (i.e., mathematics) and classic (hard) systems thinking. With this approach, operational definitions representing real world phenomena become sets of modeled parameters, allowing for snapshot observations of momentary system states. Representativeness becomes then a matter of measurement precision and model validity, aspects that enforce a detached perspective. Concerns about representativeness are thereby framed towards questions about whether parameters are valid and sufficient for describing a state, and situated meanings and dynamic characteristics are (implicitly) assumed covered by sufficiently detailed state descriptions. With mathematics as the language of science, the elapsing of time tends to become represented by sequential state transitions, progressing in steps. To simulate a continuous reality, Δt (the time-step between observed system states) is ideally made infinitesimal short. The now, the present, the current situation, what actually goes on, is then reduced to an infinitesimal transitory state described by high-precision measures. Consequently, analysis and evaluation becomes a matter of assessing past and estimating future states while the present, being no more than just a snapshot, becomes merely the last of past states. There is no now! The systemic gap is, however, fictitious. It is a consequence of using the rigidity of mathematics, often with the purpose to refute another fictitious gap.

The paradigm wars outlined in chapter 2.1 can be viewed as a fighting about how to bridge **the epistemic gap** (Figure 15.2 – A, the horizontal gap between the concrete and the abstract) that may be considered a heritage of Cartesian dualism distinguishing between mind and matter. Despite the fact that science since long has refuted this distinction, the different approaches and the disputes between the warring parties make the epistemic gap remain as a borderline entrenchment. Positivism claims that the gap is an illusion and takes stance in the concrete physical world, uses measurements of observable concrete aspects of past states and validates hypotheses by verifying predictions of future states for likewise observable and concrete aspects. In essence, positivism claims that only concrete aspects are real, implying that abstract (e.g., psychological and social) aspects are nothing but indirect or difficult to observe phenomena caused by concrete aspects. Interpretivism claims also that the gap is an illusion, but takes stance in the abstract world of ideas, uses (descriptions of) past ideas and validates hypotheses and models by assessing the relevance of gained understanding of future states. In essence, interpretivism claims that only abstract aspects

are real, at least in the sense that ideas are the only thing we can know is real, implying that concrete aspects are nothing but constructs formed by our interpretations (hence the notions of constructivism or constructionism). Both positivism and interpretivism tend, however, to adopt the mathematical analytical approach. While trying to provide rigorous proof of their claims, both approaches facilitate thereby a continued reduction of the situated present by representing phenomena by detached model-based aspects and by trying to predict future outcomes from past outcomes. Neither appear concerned about the fact that what the future states in the end will be depend on how they in practice are reached.

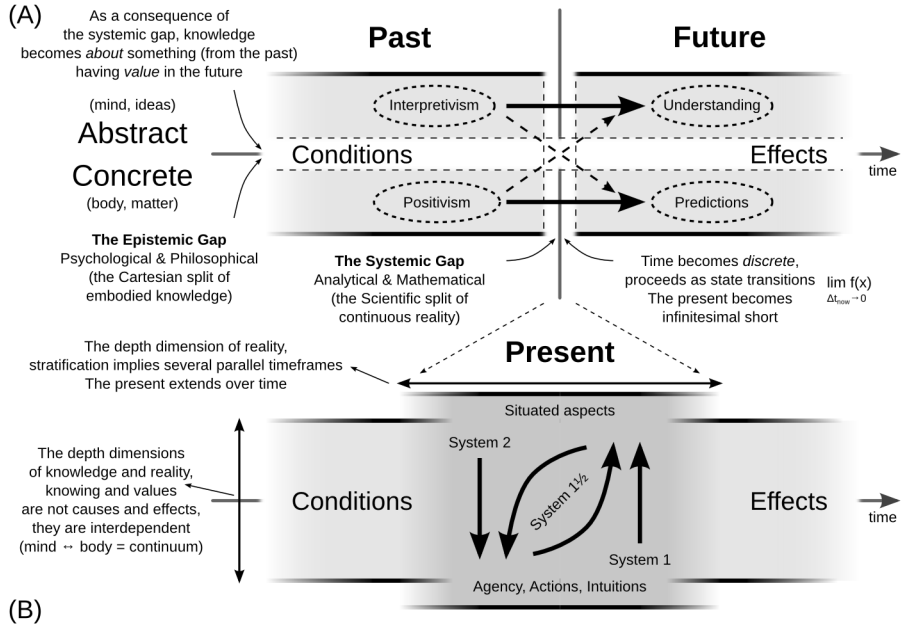


Figure 15.2: Philosophical assumptions leading to the lost situated perspective

By using systemic analysis approaches and mathematical descriptions of system states, the present does not exist, at least not more than as an infinitesimal snapshot state among others. The approach leads to what could be described as a paradox of measurements, related to the dilemma of rigor or relevance (and Heissenberg's principle of uncertainty). The more precise the measurement, the less meaning it has. In short, parametric state descriptions tend, arguably, to be deprived of rich meanings, fail to be appropriate representations of real-life dynamics, and disregard relevant contextual aspects. State descriptions seem thereby incapable of capturing all those aspects that appear to be of the greatest importance for experts, athletes, and virtuosos, aspects such as having momentum, flow, being subsumed in the present, becoming one with the tool or the instrument, and

so on. Parametric system states frames focus, especially when expressed by numbers, to detached what, while leaving out contextual why and involved how.

15.3.2 Impact on the human role

As a consequence of an absent situated present in which 'the how' takes place, aspects such as conscious decisions, actions, agency, will, intentions, and situated deliberation, but also related phenomena such as awareness, consciousness, intelligence, and so forth, become problematic. Without a rich and analyzable present, without a present in which aspects can be both causes and effects, everything becomes detached and the world becomes nothing but a deterministic machine, making any kind of situated agency appear supernatural. In fact, some scientists are rather quick to categorize certain phenomena of the involved present as paranormal, mystical, or as ridiculous ideas of supernatural human qualities (e.g., intuition and gut-feelings). Perhaps are they right. Perhaps is the world in fact a fully deterministic machine? The ultimate consequence of a completely deterministic world is, however, that human decisions and actions are nothing else than the result of conditions. Determinism implies, ultimately, that there is no such thing as a free will, that autonomy does not exist and, consequently, that responsibility is unnecessary.

Whether the world actually is deterministic, is definitely beyond the scope of this research. The present purpose is merely to stress the consequences of adopting the deterministic view to a too large extent. The daemon of Laplace is yet to be found. In the foreseeable future, there are enough uncertainties about the workings of the world, also for concrete observable aspects, to require holding people responsible for their actions. For responsibility to make sense, however, one must accept that people are autonomous individuals with free wills and conflicting interests, and that they make decisions changing how the world develops. The future is unpredictable largely because it depends on deliberate actions not yet taken. That is, for an unforeseeable future, the world is not deterministic, making what happens in the present be what matters the most. Understanding what happens in the present ought to be much more important than general predictions of what might happen in the future if the world happens to behave as our models. For contemporary science, this view seems rather challenging.

Another purpose with this philosophical discussion is to suggest that the common tendency to use discrete time when analyzing the world might be the culprit, and that the underlying kind of philosophy of science (the mechanistic worldview) might be more counterproductive for humankind than we currently think, simply because essentially we human beings are always acting dynamically in the present. When analytically reducing the

present to a snapshot state of parametric and calculative data, we reduce at the same time the importance of our own decisions and actions to nothing.

Phenomena such as intuition, gut-feelings, practical wisdom (phronesis), and the like, perhaps particularly those aspects described as associated with system 1½ properties (ch. 3.3 & 9.3) of human beings, these phenomena might become more understandable if we begin to analyze the world with a richer present. If the present is allowed to extend over some time, certain circular dependencies related to certain important situated human phenomena may become intelligible, thereby possibly removing their notions of supernaturalism (Figure 15.2 – B). The horizontal arrow in the figure depicts the present as a smooth transition from the past to the future, a transition domain in which system 1½ processes arguably takes place. However, our desire for objective facts and predictability by use of formal models, calculative numbers, mathematics, and unambiguous (i.e., monovalent, one-to-one) functional relationships, is working forcefully in favor of a discrete sense of time. Furthermore, the tendency to view time as progressing by discrete steps is arguably encouraged by our current extensive use of computers as tools for analysis.

15.3.3 Impact of computerization

Computers are built around clock-cycles altering software and hardware states in steps. Hence, for computers time is discrete and the world is made up of transitory states. Because our view of the world is shaped much by the tools we use, this characteristic of computers has, arguably, profound implications on our understanding of things. Because of discrete time, real-life quality representations are striven for by shortening the time-step as much as possible (e.g., enhancing the fidelity of moving images by increasing the frame rate), which is to mimic (i.e., simulate) continuous real-world time by using imperceptibly short time-steps. This strategy is based on the assumption that the whole is nothing but the sum of its parts (Cartesian reductionism), implying that if only the parts are sufficiently well described as in having sufficient detail, the result is indistinguishable from the whole (in fact, for reductionism, the sum is equal to the whole). Reductionism is a strategy that actually works for mathematics, as can be illustrated by the limit construct. The limit is used to calculate the value of a mathematical function where it is undefined. Metaphorically, the strategy can be described as focusing in on the problem, increasing the level of detail, until a solution is found. For example, when finding the tangent to a function f(x) by setting up $f'(x) = \lim \Delta x \to 0$ for $f'(x) = (f(x+\Delta x)–f(x))/\Delta x$. The calculated tangent, f'(x), becomes closer to the correct solution the smaller the Δx, all while the ideal case, $\Delta x = 0$ implies an undefined division by zero. The true solution is found when Δx, the present, is infinitesimal short, which logically equals being eliminated. However, for certain real-life aspects the strategy of

increasing the level of detail may in fact obstruct the understanding, such as when being unable to see the woods because of all the trees. In order to understand the meaning of the whole it is perhaps necessary to adopt the other approach, take a few steps back and look upon the entire scenery. When working with discrete time and formally defined system states, however, this approach does not work because larger time steps quickly turn the modeled representation into staggering scenery and introduce sampling problems. Real time is, arguably, qualitatively different from discrete time in a way that makes the discrete time model unable to capture the true nature of the present, regardless the size of the time-step. In particular, the smaller the time step, the poorer representation of the present, all while it is in the present that decisions are made and actions are taken. The present must therefore be allowed to extend analytically over some time, for the ability to understand truly how things happen.

Personally, I find the critical realist (CR) approach to facilitate (or at least encourage) this idea of having the present extend over some time. To begin with, the epistemic gap between ideas and matter (between mind and body, or abstract and concrete), is by CR refuted not by stating that neither concrete nor abstract aspects are more real than the other is, rather, the gap is refuted by stating that both aspects are real. Thereby, the representation of the epistemological dilemma is changed into a gap between two less sensitive aspects. The epistemological gap is moved from the emotionally touchy mind-body entrenchment to the more neutral gap between the intransitive and transitive domains of knowledge. This 'philosophical trick' is, I believe, a beautiful way to re-establish the real world as a continuous whole (Figure 15.2 – B), while still acknowledging the epistemological dilemma. The vertical arrow in the figure indicates the smooth transition between the concrete and the abstract that arguably is an effect of having to consider several strata simultaneously. Furthermore, the CR depth dimension of reality, the stratification, and the notions of rooted in and emergent from, make it necessary, I believe, always to consider several timeframes simultaneously. In practice, stratification has therefore the effect of opening up the present to occupy necessarily more than an infinitesimal amount of time (represented by the horizontal arrow in the figure), simply because different strata require different timeframes. Within the richer present, it is possible to consider system 1 and system 2 as simultaneously affecting each other, effectively forming a system 1½.

16 Discussion

16.1 The logic of discovery

The purpose of this research was to explore usefulness critically and based on the gained understanding explain what makes technological systems useful. During the first research phase, usefulness was explored by problematization and by a breaking down of the concept into basic questions, about what should be useful, when and where it should be useful, and for whom it should be useful. A sociotechnical worldview was presented (ch. 5.2), placing the technology of concern in a physical and social context with possibly distinct work and effects domains. The worldview model outlines different social and physical systems as well as different categories of involved human beings connected by the technological system of concern. There were also conceptual frameworks presented, describing technological properties in general (ch. 5.3) and computerized properties in particular (ch. 5.4). This mainly ontological exploration was then followed by the most important question of all, the question why (ch. 6). For what purpose should the technology be useful? However, for every technological system that is more than a trivial system-component with a bounded overall influence on effects, the question why opens up to a system of interwoven social and technological issues, virtually unlimited in complexity, inescapably ending in ethical dilemmas and moral standpoints reaching far beyond the scope of the present research.

For the present purpose it suffices to conclude that usefulness in the end is a subjective value, although mostly associated with certain objective aspects and concrete conditions, but ultimately a subjective judgment made by human beings, and to realize that this subjectivity in fact applies already to basic system components however deterministic and well-defined their effects and purposes may appear. In the end, what matters for judging the usefulness of a technological system is how actually occurring effects align with truly desired effects. Arguably, both predicted and situated effects depend on how they come to be. The effects that actually occur from using the technology in a situated reality are ultimately determined by how system properties are applied, they are not determined unambiguously by systems design. Possible discrepancies between predicted effects according to design intentions and actual effects come in part from the fact that even the physical world is not completely predictable, but perhaps they come much more from

the fact that local purposes and value judgments are inherently unpredictable. They are unpredictable because future purposes and values are to some extent based on deliberate actions not yet taken. That is, purposes and values are situated phenomena. The effects that in reality are desired depend on subjective values based on personal interpretations of effects actually occurring, and on how they have come to occur. Furthermore, they depend on culturally dependent individual (or group) real-life experiences and expectancies of what is supposed to happen. This fact creates a circular dependency between what actually will be desired and how effects occur in practice. Truly desired effects are determined by values judged by people really involved in the situation where the effects occur, not by values defined by detached models and calculative measures.

How technology can be used and what effects that are physically possible to achieve are limited, however, by the physical properties of the system that in turn are determined by its design. Construction and design of technology is thereby essentially about solving a complex puzzle of clashing purpose-property relations that for physical aspects must be settled before the concrete product actually can be built (e.g., it is impossible to build something with larger inside physical dimensions than outside dimensions, no matter how desirable it would be). The possibility of effects, determined by systems design, both intended (i.e., predicted) and emerging possibilities of effects (e.g., from a creative discovery of alternative purposes), was used to define the concept of utility (ch. 6.2), indicating a potential to become useful. The potential for useful effects is at large determined by physical aspects, and the physical properties of a system are always present, impossible to disregard. Clashing physical properties may therefore remain as design compromises, and some opposing physical properties tend to neutralize each other. For example, the possible acceleration with a strong engine is hampered by the heavy chassis required to hold it in place. Because of this neutralization, one combined physical property emerges, an average. Purposes, on the other hand, are possible to disregard. Clashing purposes may therefore remain with full implications, even after the design is completed. During design, properties are selected according to purposes that are considered adding to a predicted desired functionality, while other purposes are overruled out of necessity (e.g., due to limited development resources or incommensurable aspects), disregarded from being out of focus, or not considered simply because they are unknown. These other, possibly evil, purposes-property dyads may later be revived or freshly discovered when the system is used, which is what makes purpose-clashes remain and the value of possible system effects ambiguous. There is a tension within the concept of utility. On the one hand, a bounded physical tension in the form of design compromises, and on the other hand, a virtually unbounded (social) tension between purposes.

During design, one methodological source of unpredictability is a mismatch between the character of the models used for predictions and the character of the predicted reality. This kind of mismatch is probably why purposes are much more difficult to predict than physical conditions. Models used for predicting desired concrete effects, thereby implicitly assuming or predicting future purposes, are often developed by use of systemic decomposition and classic (hard) systems thinking, an approach that facilitates rather accurate predictions of physical aspects. Real-life purposes, on the other hand, depend on social aspects not necessarily having the deterministic character of physical systems thereby making predictions by use of such hard system models largely irrelevant. Or as Taleb (2010) phrased it, while the physical world is the self-moderating mediocristan, the social world is the self-generating country of extremistan, implying that only the former might be appropriately represented by the ubiquitous normal distribution of probability (i.e., the Gaussian bell-curve) on which many predictions rely. Social systems are inherently unpredictable as they depend on conscious decisions based on arbitrary local values that feed on each other, which is quite in contrast with physical systems that depend on natural factors generally centered round a mean value with a calculable variability.

The social tension within the character of utility is considered to represent the ethical dilemmas connected with the concept of utility and with related value judgments of potential and realized effects. The notion of a tension is explicitly used to indicate that for a certain technological system, it is not the question of having one kind of utility or the other, the character of utility is a rich concept consisting of several, probably conflicting, aspects that must be considered altogether. Usefulness and utility were then identified as different concepts, although with subjective values and use (i.e., actual and possible application) as the common ground. Utility was defined as concerning the value of conditions for use while situated usefulness was defined as the value of using these conditions in a situation (ch. 6 & 8.3).

Grounded in overall experiences (ch. 3, perhaps particularly chapters 3.4 & 3.5) and in studies of overarching external theories such as ethics, philosophy, and general systems thinking (ch. 5 & 6), the initial problematization of the concept of usefulness led to an identification of the crucial difference between the involved and the detached perspective, or that between a calculative and a generative approach, between stereotypical and individual behaviors, between predefined and contextual meanings, between model-based and situated values, between knowing-that and knowing-how, between theoretical and practical wisdom, and so forth (ch. 2.1, perhaps particularly 2.1.4 & 2.1.5). The essential problem with the contemporary view of usefulness had thereby been sifted out as being the prevailing tendency to omit situated aspects, a tendency arguably coming from an exaggerated focus on predictability and objectivity.

16.2 The core insight – true usefulness is situated

Usefulness comes from valuing subjectively how effects come to be, a statement actually comprising the core of the insight and constitute thereby the most important research message to convey. The message is essentially that 'the how' often matters more than 'the what'. How determines both what effects that actually will occur and how the effects will be interpreted. The value is determined by 'the how' much because we human beings understand and interpret things not only by observing conditions and results but also from how we take part in producing these results. The argument is not primarily that involved interpretations always are better than detached. Some of our biases are certainly unfortunate, and some heuristics unfavorable, which may imply erroneous interpretations. We do have reflexes and instincts, sometimes leading to non-optimal actions, at least if actions and results are analyzed from a position endowed with the larger picture that most often is accessible only with hindsight. Rather, the most important aspect of the argument is about the difference in character between involved and detached values (the bottom part of Figure 16.1).

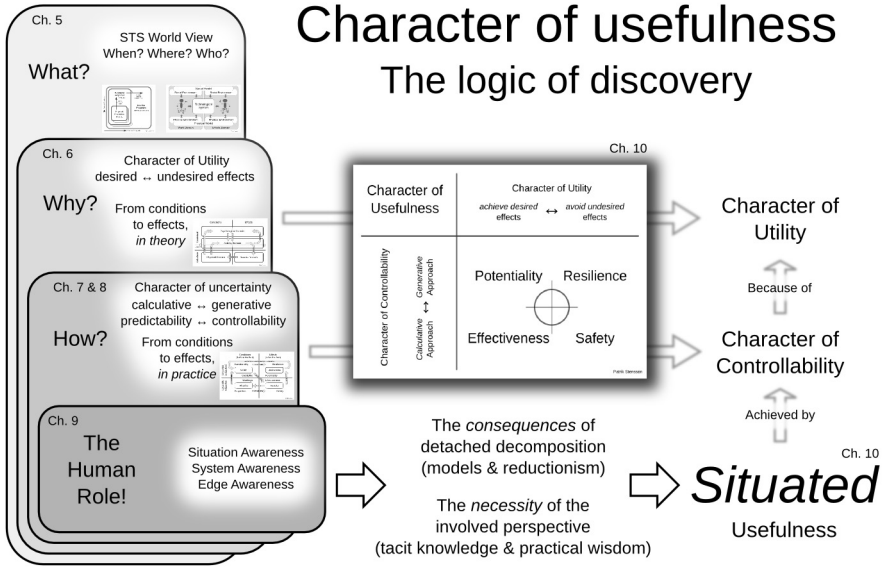


Figure 16.1: Character of usefulness, the logic of discovery

If we merely observe results from a distance, either spatial, temporal, or some kind of emotional distance that often is the result from being out-of-the-loop, we tend to attribute only detached values (i.e., formal measures) to outcomes while involved and subjective values tend to become disregarded

(ch. 7, 8 & the philosophical implications discussed in ch. 15.3). The core insight is about the insufficiency of detached values. Another way to put it is to state that it is not always appropriate to assume that objectively correct values are relevant. The first problem lies in the assumption that values actually can be objectively correct, the second in assuming that what is logically (i.e., objectively) correct must also be relevant. Arguably, the reason why usefulness is defined by supposedly attainable objectively correct values is because the kind of reasoning used while studying physics is applied also to social aspects that define values, hence the notion of technical rationality. Rigorously calculated aspects are doubtlessly often useful, especially in the sense that without them a system often becomes useless. However, for socially and contextually defined values relevance comes often before rigor, regardless the calculative truthfulness (i.e., logical consistency). There is more to situated usefulness than objective effectiveness and calculative safety. To be relevant and considered useful in a situated sense a system must also be resilient and facilitate potentiality, aspects that depend on local conditions and contextual values. Situated usefulness is more about how the system can be used than what it is intended to achieve. This insight called for further scrutiny of how effects from using technological systems in fact come to be. In chapter 6.3, a model of how effects come to be, in theory, was presented. For an ability to explore in more detail how effects come to be, in practice, a complementary model was presented in chapter 8.3.

In the end, how effects come to be is always the result of human activities, regardless the level of automation. Every technological system is designed, built, bought, applied (i.e., used and, perhaps, controlled) and maintained (e.g., serviced) by human beings. There is no such thing as autonomous technology making decisions and performing activities for own reasons based on autonomously conceived purposes, and my position is that there should never be anything like that (cf. ch. 15.2). How effects come to be depend thereby on the human role, which is a role that we as autonomous human beings define for ourselves. The main critique of the contemporary view of usefulness, thus a critique directly addressing the second research purpose, is that the prevailing view of usefulness tends to have us define our role too much according to formal models with detached representations of values. We thereby define our own role in terms of what can be expected by rational machines, a role in which computers excel. The most pronounced critique of the contemporary view of usefulness is the implicit flight from responsibility that follows from such exaggerated focus on detached aspects. Models and technology cannot take responsibility for effects affecting human beings because they have no life at stake to which responsibility ultimately relates, making responsibility a role from which we human beings cannot escape.

However, for the ability to understand fully what responsibility truly means in a real-life situation, we must apparently be sufficiently involved to have our intuitive system 1 take active part in what happens (ch. 9). Without such low-level involvement, we tend to refrain from applying higher intuitive human concerns such as empathy to our activities and regress to attribute merely stereotypical meanings, which are meanings similar to the model-based values that govern the behavior of automated systems. Without intuitive concerns, we tend arguably also to lack the sense of urgency to make an effort and thereby we refrain from using the rational system 2 to its full potential. This argued connection between low-level involvement and higher-level reasoning is the rationale for discussing the semi-rational system 1½. The conclusion to be drawn from the discussion about system 1½ is that low-level involvement for human beings is an end in itself. That is, to have our intuitive system actually take part, for actually having an ability to enact the expected responsibility, systems must be designed to provide means for situated system control, which implies means for understanding the system such that we can develop incentives to control it. All the contributed frameworks and models are based on the assumption that the human role is to make systems useful and take responsibility for system effects by use of situated controllability. The theoretical contribution is therefore focusing on the characteristics of the involved perspective, the importance of tacit knowledge, the significance of skills and the kind of rich understanding that only can be achieved by having the own body involved, as well as the necessity of experience and practical wisdom.

16.3 Situated usefulness explained

Up until now, this concluding discussion has essentially been about the left hand side of Figure 16.1, the logic behind discovering underlying aspects and mechanisms of situated usefulness, a process driven by problematization and a breaking down into aspects leading to identification of issues associated with usefulness of technological systems. The following explanation of situated usefulness is about the right hand side of the figure, the logic behind putting it all together and reasons for why these underlying mechanisms are relevant. Situated usefulness as a concept, its general character, and the model describing the character of usefulness (repeated in the middle of Figure 16.1), was explicitly explained in chapter 10.2.

If the premise can be accepted that the human component is crucial for making the usefulness of technological systems situated, the character of the human being in the role of an involved system controller becomes the core issue (ch. 9.1). For the role as a system controller, the concept of situation awareness was found intuitively important, but still considered an insufficient as well as a problematic concept, partly due to inconsistent

interpretations. In particular, when used as a cause of undesired outcomes (e.g., lack of situation awareness as the cause of an accident) it fails to acknowledge the active side and the deliberate aspect of the human component. Instead, it promotes the view of human beings as passive information processors and executers of logical programs (similar to computers) and thus the view of outcomes as the result of information processing quality. Such interpretations of situation awareness appear to disregard the impact of human agency and the situated development of incentives and intentions. The human controller may, for example, be perfectly aware of what is going on but still deliberately act in a way that causes an accident (i.e., having an evil purpose). To begin with, as a remedy for the unfavorable passive view, situation awareness was complemented by the concept of system awareness (ch. 9.2), aiming to sort out some confusions associated with what possibly relevant awareness is about. Then the title concept, edge awareness, was introduced (ch. 9.3). The purpose of the concept of edge awareness is, in combination with system awareness and the more specific version of situation awareness, to better depict the active side of the human component by capturing relevant conditions for agency, while still not aiming to represent agency itself. Edge awareness does not cause things to happen, actions do.

Edge awareness is considered a phenomenon residing in the borderland between the intuitive system 1 and the rational system 2 (Stanovich and West 2000, Kahneman 2003, 2011), a product of both the mental systems working in conjunction as one system 1½ (ch. 3.3 & 9.3, and Figure 15.2 – B). Edge awareness is presented as a semi-rational intuitive awareness, fundamentally grounded in hands-on experiences and personal skills that mostly are of a tacit character, governed only in part by rational reasoning that, however, is based on the intuitive understanding that comes from experiences and skills. Edge awareness is about knowing-how to control a system such that real performance edges can be balanced on, edges that preferably are known objects of knowing-that, and perhaps most important, it is about intuitively knowing the limits of the own ability to do so. For the possibility to maintain such intuitive knowledge, a continuous and active taking part in control activities for the system of concern is required, which in turn requires the system having been explicitly designed to facilitate a generative character of controllability.

The character of controllability in the case of use determines how the human component can take part in the outcomes of systems operation, thereby also shaping interpretations of effects and judgments of outcome values. It is therefore of great interest to understand what determines the character of controllability of a designed technological system, which, of course, is determined much by the approach in the case of design to controllability in the case of use (i.e., designers' view of the human role as a system controller). Because the resulting conditions for controllability are

determined by systems design, a system will obviously have the principal character of controllability that the designers find appropriate. The tension within the character of controllability was thereby concluded as mainly occurring between the case of design and the case of use. For the ability to know what to build, a calculative approach is required, which, however, risk creating technological systems with a calculative character of controllability, if predicted usage scenarios and purposes of use are taken too seriously. The design dilemma is in principle the same as the safety preparation dilemma (discussed at length when interpreting the Fukushima case in ch. 12.4.2). While too rigid preparations create brittle or false safety that risks create disasters, such preparations create also stereotypical or illusory usefulness that risks become irrelevant effectiveness.

Because design is a creative process that relies on people communicating, the greatest threat against generative controllability is a reduced richness of concepts required for designers and evaluators to understand the involved perspective. The more detached the common understanding becomes about the meaning of concepts with involved connotations, such as usefulness and utility, functionality and capability, intelligence, awareness, consciousness, responsibility and autonomy, the more these concepts become reduced to mean merely technological properties, the less the involved perspective will be part of the design equation. When the calculative approach to controllability spills over and becomes a calculative approach to judging values as well, the vicious circle culture (R3) is an accomplished fact. The map has then become more important than the territory.

16.4 Validity of contributed models and frameworks

Critical realist (CR) philosophical assumptions were characterized as ontologically bold and epistemologically cautious (ch. 2.1.1), which in a sense matches the ambition of the present research. While rather boldly presenting plausible descriptions of the nature of situated usefulness (i.e., the ontology of the studied phenomenon), the initial claims of rigorous validity for the developed models and frameworks are rather modest. Testing for such validity is essentially left for future research. However, the methodology used grounds for a certain level of plausibility of the models and frameworks generated through the theorizing process in phase one. This level of plausibility is considered sufficient in light of the overall purpose of this research that was to initiate a social transformation of the contemporary view of usefulness identified as unfavorable for humankind, a purpose for which plausibility and relevance is of overarching importance compared to rigorous predictions. However, to indicate further validity, phase two added empirical data on which the presented models and frameworks were applied. The results from scrutinizing the cases were interpreted qualitatively.

Central to CR methodology is the concept of retroduction, which implies developing plausible explanations of the studied phenomenon by use of deduction and induction in parallel (i.e., similar to pragmatist abduction), as well as developing reasons for why the explanations are plausible. The frameworks and models from phase one serves both purposes. Models categorized under R2 explaining usefulness, in particular R2.2, constitute the contributed plausible explanation of situated usefulness thus answering RQ1, and models categorized under R1 provide reasons for why R2 is plausible. The core insight, the tendency to analytically remove the involved perspective and thereby facilitate a reduction of concept richness that in the end implies an escape from responsibility is explained by R3, the vicious circle culture model. Overall, the contributed explanation of situated usefulness (R2) and the supporting models and frameworks (R1 & R3) were found to facilitate novel and presumably relevant descriptions of essential aspects for all the empirical cases. The plausibility of the explanations is thereby considered further grounded. As the plausibility nevertheless still rely rather much on the reasons provided by the supporting frameworks, a discussion about how all these frameworks are linked together follows (outlined by Figure 16.2), to allow for further assessment of validity.

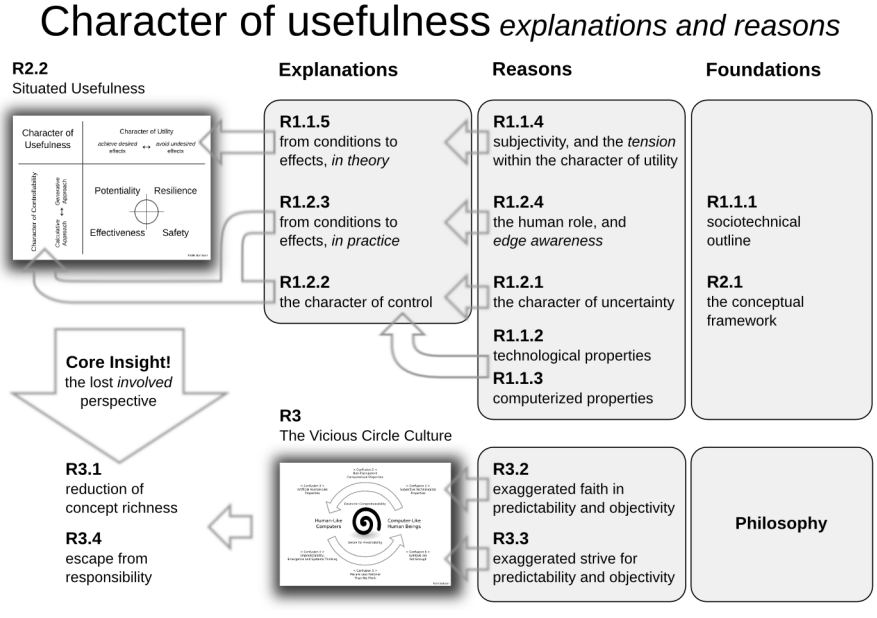


Figure 16.2: Contributed models and frameworks, explanations, reasons for plausibility, and foundations for reasoning

The model of situated usefulness (ch. 10, in particular Figure 10.2, R2.2, p. 160) explains usefulness by depicting **the character of usefulness** as a rich

characteristic described by four keywords, effectiveness, safety, potentiality, and resilience, arranged in a two-by-two matrix with tensions as its dimensions. On the one hand, there is the dimension of what effects that will or are intended to occur. On the other hand, there is the dimension of how these effects will come to be. What effect that will occur (in principle) is explained by the map of how conditions become effects in theory (ch. 6.3, R1.1.5), showing that, ultimately, it depends on human activities regardless the level of automation. The ultimate dependency on human beings is to begin with practical because truly autonomous technology applying its properties for its own reasons does not exist. Furthermore, the human role is inescapable because the purpose when applying technological properties in a situation is what determines the value of whatever effects that occur, and purposes are defined by autonomous human beings. The two overarching categories of principal purposes were concluded to be either the achieving of desired effects (i.e., effectiveness and potentiality) or the avoiding of undesired effects (i.e., safety and resilience). Purposes may, however, clash and effects may be valued differently by different people, a fact that further enforces the necessity of the human role as a system controller, taking responsibility for system effects. The tension within the character of utility (ch. 6.2, R1.1.4) is the reason why the map of how conditions become effects is centered on the human role.

How system effects will come to be is explained by the map of how conditions become effects in practice (ch. 8.3, R1.2.3), showing that the difference between calculative achieving of predicted model-based effects and generative achieving of situated effects is related to the character of human involvement. The title concept framework, the system, situation and edge awareness distinction (ch. 9.3, R1.2.4), provides reasons for why human involvement is essential. We human beings must simply be involved to become aware, we must be aware to maintain control, and we must be in control in order to make effects contextually relevant and useful in the present situation. How system effects will come to be is also explained by the character of control (ch. 8.2, R1.2.2), determined by the properties of the technological system of concern, the system to be controlled by the human operator. The two extremes of control characteristics are described as either a calculative character of control focusing on achieving predictable effects according to well-defined models (i.e., effectiveness and safety, detached values) or a generative character of control focusing on achieving contextually relevant effects according to situated aspects (i.e., potentiality and resilience, involved values). One reason for the necessity of situated control is the inescapable presence of uncertainties, described by the character of uncertainty (ch. 7.2, R1.2.1). Another reason is that, in the end, the resulting character of control for the designed system of concern depends on its technological properties (ch. 5.3, R1.1.2) and certain characteristic properties of computerized systems (ch. 5.4, R1.1.3) are inherently

problematic for human beings to comprehend, which thereby becomes a reason for why the character of control may become unfavorably calculative.

The remaining two frameworks from phase one, the overall sociotechnical outline (ch. 5.2, R1.1.1) and the conceptual framework connecting most of the concepts discussed (ch. 10.1, R2.1), works as a foundation for further analysis. Now it is perhaps once again necessary to stress that Figure 16.2 does not depict causal relationships or logical chains for proof of validity, which, if it did, would imply a circular dependency where the present research would be validated by the theory it developed. The methodology used relies on plausible explanations and reasons for why explanations are plausible, forming a discourse for which concept definitions and structural models are used to clarify meanings. In this context, the label 'foundations' should not be interpreted as axioms but as further reasons for plausibility of explanations, or as additional explanations for plausibility of reasons. When an indisputable and unambiguous objective truth does not exist (e.g., a correct answer to the question of what usefulness actually is), in the end it all comes down to conveying convincingly a message about what usefulness might be. The figure depicts therefore a rhetoric chain of arguments. However, the considered clarity that comes from having these rhetoric arguments packaged as depicted by Figure 16.2 does, arguably, provide ample scope for more rigorous exploration of issues and relations inherent in these descriptions. The main contribution is that now such explorations are endowed with an adjusted taxonomy and a set of models and frameworks that include the involved perspective.

17 Conclusions

The initial argument was that the heritage from the industrial era, the paradigm of technical rationality, prevails and governs the contemporary view of technological usefulness. Essentially, the prevailing view is a mechanistic and deterministic view of the world in which desired effects and safe system operation procedures are considered derivable beforehand by use of sufficiently detailed models and capable computational powers. Within such a paradigm, implicitly the ideal behavior of people in general and for system operators in particular is to act predictably, like rational machines. Because of this view, system control becomes abstracted and automatized according to formal models since explicit model-based values are assumed objectively correct and relevant. This kind of calculative approach is believed to ensure both safe and effective system operation. The main problem with the calculative approach is that the world is often not predictable. Predetermined system operation is not safe and predefined effects are not necessarily desired, more than in a superficial and general sense. Hence, model-based safety is sometimes false safety and model-based effectiveness can sometimes be illusory effectiveness. True safety and truly desired effects emerge in the situated reality that models are merely stereotypical representations of, and real safety as well as truly desired effects are both required before a technological system can be considered truly useful. True usefulness is situated and models are detached. Model-based usefulness is therefore often insufficient.

Besides the following five conclusions directly related to the research purpose and objective, the concept of usefulness was explored by a methodological approach inspired by critical realism, which is a method shown to provide new insights. Critical realism was found very helpful for the analysis of situated technological usefulness, presumably because usefulness is a cross-domain phenomenon, a socially situated phenomenon about something from the domain of the natural sciences. The method used in this thesis may therefore be applicable to scrutiny of other similar concepts and phenomena as well, and constitute thereby a scientific contribution of its own. In addition, the critical realist philosophy was found useful for explaining scientific aspects of the lost situated perspective, a fact that might constitute a contribution to the philosophy of critical realism.

The first conclusion is that the contemporary view of usefulness is calculative, a conclusion that **is directly related to the main research purpose** to scrutinize and reflect on the contemporary view of usefulness. After having studied four incidents in two ultra-safe domains and applied qualitatively the contributed definition of usefulness, the initial rhetoric argument about the contemporary view still being governed by industrial age technical rationality has been strengthened, both theoretically and empirically. The contemporary view of usefulness was found to lack situated aspects to a worrying degree within both the air transport sector and the nuclear power production business. The present research supports thereby other work within the human factors and safety science domains, especially the kind of work that focus also on the human contribution rather than mainly on human error (e.g., Hollnagel *et al.* 2006, Vicente 2006, Reason 2008). While human factors, arguably, tend to focus on shortcomings such as threats to safety as a consequence of lacking resilience, which can be interpreted as a lack of situated safety, this research suggests that there is an analogous threat to effectiveness as a consequence of lacking potentiality, that is, lack of situated effectiveness. As such, the present theoretical contribution should provide interesting new grounds to explore as well as means to do so.

The second conclusion is that situated controllability is an end in itself, a conclusion that **is an essential part of the answer to both RQ2 and RQ3**. The theoretical contribution assists in keeping the situated perspective in mind and in explicating the inescapable human role as a responsible meaning maker, a role that requires involvement because meanings and responsibility largely are tacit and intuitive aspects not possible to understand without taking active part in what happens. Taking responsibility is an essential part of the answer to RQ2, another essential part is how that is made possible or how it is counteracted by system designs, which thereby is an essential part of the answer to RQ3 as well. System controllability is required because the world is unpredictable, but also because without actively controlling the systems, we are unable to make insightful and relevant decisions about desired effects, and appropriately judge the values of these effects. **The theoretical contribution answers RQ1, thereby meeting the research objective** to contribute an alternative definition of usefulness. The theoretical contribution relates also to the overall research purpose to scrutinize the contemporary view of usefulness because the contributed models and frameworks help distinguish what the critiqued view lack.

The third conclusion is that the premises for the calculative view remains and seems further strengthened. After having studied four incidents from two timeframes as well as a few future scenarios, and after having applied qualitatively the reasons for why the contributed alternative definition of usefulness is plausible, it appears that insights presumed necessary to refute

the calculative view and slow down the calculative development is missing. In the past cases, issues clearly identifiable as coming from a calculative approach to controllability have become worse. In the future scenarios, the situated perspective was mostly absent, either by equating the possibility to develop certain technologies with future usefulness or from focusing solely on achieving predicted effects according to model-based values. The mechanistic view of the human role, as a system component expected to strive towards achieving predicted effects, is evident throughout all cases and scenarios.

The fourth conclusion is that a vicious circle culture prevails, a culture where the richness of concepts essential for explicating the situated perspective becomes reduced. Concepts traditionally used exclusively for human qualities are being reduced to have detached meanings applicable to technological properties. This **is a conclusion related to the second research purpose**, to reflect on the impact that the contemporary view of usefulness has on human autonomy. The culture appears to thrive on two mutually supporting aspects. First, there is an exaggerated faith in objectivity, driven by the natural desire for comprehensibility, a desire that makes us mistake the map for reality, consider artificial systems equal to ourselves, and define usefulness according to calculative values. Second, there is an exaggerated faith in predictability, driven by the natural desire for an absolute truth and absolute safety, a desire that makes us enforce predictability on human behavior, escape from responsibility, downplay experience, and lose focus on the situated perspective.

The fifth conclusion is, consequently, that the theoretical contribution and the conclusions drawn from the case studies add valuable aspects for scrutinizing technological usefulness. More specifically, the threat to situated effectiveness that comes from having a calculative approach to controllability is perhaps even more serious than the threat to situated safety. While accidents are tangible and thereby identifiable, even if they are rare, the lack of potentiality is much more difficult to identify, especially when it becomes a non-issue because value concepts have been deprived of the situated perspective. Such problems become philosophical and tend to be dismissed as irrelevant in a world focusing on concrete model-based values. The aspects that the present theoretical contribution explicates are therefore of utmost importance. They are aspects necessary for understanding what is required when developing useful systems, for understanding what is needed to achieve situated usefulness, as well as for understanding how to avoid having humankind blindfolded by technology and merely varying randomly around a mean predetermined as desired by a simplified model of reality. The theoretical contribution is important for understanding what is required for having technology support autonomous human beings. The quest for edge awareness is an end in itself and, apparently, a lesson not yet learned!

17.1 Future research directions

From my own perspective, two major future research directions are identified as especially interesting. The first is to continue along the path that made me select the present research topic to begin with, and start sorting out consequences of thesis results, perhaps particularly for the military domain. The current development in the world regarding international security relations, towards a more fragmented distribution of powers with less firm connections to nations, makes sheer technological performance superiority less certain to be relevant. For performance duels to be relevant, it is required that the structure of future competitions is fairly well known. The development today, however, seems to be otherwise. Consequently, the social dimension with norms, values, desires, and intentions, becomes increasingly important for judging technological usefulness. Military-technology as a scientific topic, focusing on usefulness of technological systems in military situations that include difficult social judgments, is therefore in my view a forerunner for studies of technology, a kind of scrutiny that can be followed by studies of civilian technologies as well. The normative and will-based aspects that are essential for politics and military strategic decisions appear largely missing among contemporary technology developers that instead appear governed only by business aspects. Trouble is that the market is not an ethical decision maker, yet we often refer to the free market as a guarantee for sensible products. Perhaps is the ultra-extreme military environment necessary for overruling economic aspects and for truly understanding the implications of not being in control of the technology that we use? The human role as an involved system operator and the relation to autonomy and responsibility are aspects of usefulness that feel important to continue pursuing.

The second path is somewhat related to the first, but more specifically about the human being in the role as an autonomous and responsible decision maker. It appears there is much left to understand about human beings and about how we actually make decisions. The contemporary view of usefulness governed by technical rationality, for example, is quite easily connected with classic decision-making theories that clearly use detached models as value references (e.g., coherence theory of truth and mathematics as the language of science). Naturalistic decision-making (e.g., Klein 2008) and dynamic decision-making (e.g., Brehmer 1992) promote a different view. The present theoretical contribution with its notion of system 1½ as an important aspect of situated in-the-loop control decisions, together with the edge awareness construct, should provide interesting input to this area.

18 Summary in Swedish

Jakten på medvetenhet om gränser och marginaler
Kunskap som vi ännu verkar sakna

en avhandling

om avancerade tekniska systems praktiska och situerade nytta
i en ständigt föränderlig värld med ofrånkomliga osäkerheter
och motstridiga intressen

Det spelar ingen roll om du tycker om att hålla dig med goda säkerhetsmarginaler, om du gillar att testa gränser, eller om du hamnar i en situation där du måste besegra en motståndare genom att vara skickligare i att hålla dig på rätt sida om någon sorts gräns. Oavsett situation så är det mycket viktigt att 'ha koll' på var gränserna går. Om du inte har koll på marginalerna kan du lätt snubbla över kanten och falla hjälplöst ned i vad det nu är som finns på andra sidan. Det spelar heller ingen roll om du är barn eller vuxen, om du är professionellt engagerad eller agerar individuellt. Idag inkluderar i princip allt vi gör avancerade tekniska system och kanter finns överallt. Problem i interaktionen mellan människa och system samt bristande '*edge awareness*', kantmedvetenhet, eller medvetenhet om var olika gränser går, är därför av betydelse för alla!

Avhandlingen handlar om människor i ett högteknologiskt samhälle och om vår roll i förhållande till den teknik vi utvecklar och använder, vilket är en mycket mer komplex företeelse än individer som interagerar med tekniska prylar. Hur vi ser på vår egen roll relativt den teknik vi använder påverkar också vår syn på teknikens nytta, vilket i sin tur påverkar hur vi ser på vår egen roll när vi nyttjar tekniken, ett cykliskt förhållande på gott och ont. Denna avhandling granskar kritiskt vår roll som användare av tekniska system, i syfte att förbättra vår förståelse av nyttan med modern teknik.

Den traditionella och ännu dominerande synen på nytta med tekniska system kan, i alla fall som utgångspunkt för en diskussion, sägas vara baserad på en teknisk rationalitet med rötter i den industriella tidsepoken. Idag förordas inom detta paradigm en modellbaserad syn på betydelser och värderingar, en syn som ofta och i alltför stor utsträckning bortser från situerade och kontextuella aspekter. Jag påstår i denna avhandling att denna

överdrivna fokusering på förutsägbarhet enligt deterministiska modeller och teknisk rationalitet gör att tekniska system, som förutsätts vara både säkra och effektiva, i stället utformas så att en konsekvens är att de skapar falsk säkerhet och inbillad effektivitet som därmed blir bedräglig, vilket resulterar i en typ av nytta som både är suboptimal och kontraproduktiv från ett mänskligt frihetsperspektiv.

Världen är komplex och i stor utsträckning oförutsägbar, speciellt när det gäller bedömning av värderingar, tolkning av betydelser och skapande av incitament. Verklig nytta kräver därför att effekterna av de system vi använder kan anpassas till de lokala förhållanden och sammanhang som gäller i de situationer då systemen verkligen används. Vissa situationer är kanske mer oförutsägbara än andra, till exempel militära situationer som beror på vad intelligenta motståndare tar sig för. Min bakgrund som stridspilot och Human Factors-forskare kan därför vara en speciellt intressant utgångspunkt eftersom den har gett mig personliga erfarenheter både av att använda och analysera användningen av avancerade tekniska system i extrema situationer där gränsen mellan succé och fatalt fiasko ofta är skarp och omedelbar, vilket därför motiverar att tala om en kant. Denna praktiska och situerade nytta som är nödvändig i sådana extrema situationer utgör modell för en uppdaterad syn på nytta som visar sig tillämpbar på mycket mer än det specifika fallet med militär flygverksamhet. Kort sammanfattat så är syfte och mål med denna avhandling följande:

- att kritiskt granska den nuvarande synen på begreppet nytta (av tekniska system)
- att reflektera över den modellbaserade synen på nytta och vad den har för konsekvenser för synen på mänsklig autonomi
- att bidra med en alternativ definition av nytta som tar hänsyn till:
  - det situerade och det kontextuella
  - det oförutsägbara

Avhandlingen problematiserar begreppet nytta, delvis genom att driva frågor till sin spets. Utgångspunkten är den nuvarande tekniskt rationella synen på nytta där väsentliga aspekter glöms bort av det enkla skälet att de inte finns med i de modeller som används för att definiera nytta. För att granska nytta av tekniska system, vilket är ett socialt fenomen om fysiska ting, så är både tolkande och positivistiska ansatser otillräckliga. Båda synsätten är nödvändiga. Kritisk Realism stödjer båda synsätten, möjliggör att kombinera aspekter från olika paradigm och bidrar med metodologiska riktlinjer för att göra detta. Med filosofiska antaganden från Kritisk Realism kan gränserna för teknisk rationalitet överskridas och det blir möjligt att åstadkomma en definition av nytta som även tar hänsyn till det situerade, det kontextuella och det oförutsägbara. Målsättningen är att denna definition ska bidra till en social förändring.

Flera begrepp relaterade till nytta granskas, omdefinieras, eller kompletteras, begrepp såsom förutsägbarhet, kontrollerbarhet, effektivitet och säkerhet. Underliggande faktorer och mekanismer utforskas och inbyggda spänningar identifieras, vilket resulterar i ett teoretiskt bidrag i form av modeller och ramverk som anses förklara vad verklig nytta är. Möjlighet (potentiality) föreslås komplettera effektivitet (effectiveness) på samma sätt som resiliens (resilience) kompletterar säkerhet (safety). Situerad nytta definieras sedan genom att använda dessa fyra begrepp. Fenomenet situationsmedvetenhet (Situation Awareness) granskas också, samt kompletteras med begreppen systemmedvetenhet (System Awareness) och titelbegreppet kantmedvetenhet (Edge Awareness).

Fyra verkliga fall, två flygolyckor (SAS SK751, Gottröra, den 27:e december 1991 och Air France AF447 som störtade i Atlanten, den 1:a juni 2009), en kärnkraftsincident (Forsmark den 25:e juli 2006) och en kärnkraftskatastrof (Fukushima, den 11:e mars 2011), samt tre framtidsscenarier, utgör det empiriska bidraget. Analyserna visar att avhandlingens teoretiska bidrag medger alternativa eller utökade förklaringar av alla fyra fallen, samt att framtidsscenarierna saknar tankar om situerad nytta. Implikationer från denna forskning kretsar kring den mänskliga rollen och våra ansvarsområden i förhållande till de tekniska system vi använder, samt kring betydelsen av de koncept som definierar denna roll.

Vi är situerade varelser. Vår roll är att vara personligt involverad och ansvarig för vad som händer, vilket är en roll som kräver insikt och kontrollmöjligheter. Automation och vissa karakteristiska egenskaper hos datorer har dock en tendens att begränsa kontrollmöjligheterna samt motverka både medvetenhet om systemens egenskaper och om de situationer som uppstår eftersom systemen själva i stor utsträckning presenterar situationerna för oss, vilket då också motverkar medvetenhet om prestandagränser, eller prestandakanter. Dessa uppenbart ofördelaktiga konsekvenser av modern teknik är förmodligen oftast inte avsiktliga, utan uppstår kanske främst på grund av att vi ännu inte har dragit nödvändiga erfarenheter från användandet av sådana system. Det kan också bero på att vi i allt för stor utsträckning analyserar systemens funktion och värderar effekterna med hjälp av modeller som saknar de aspekter som påvisar behovet av mänskliga insikter och kontrollerbarhet. Det ständigt ökande inflytandet från datoriserade tekniska system på vad vi gör och hur vi agerar gör därmed jakten på kantmedvetenhet viktigare än någonsin.

Bibliography

Ackoff, R.L., 1979a. The Future of Operational Research is Past. *The Journal of the Operational Research Society*, 30 (2), 93–104.

Ackoff, R.L., 1979b. Resurrecting the Future of Operational Research. *The Journal of the Operational Research Society*, 30 (3), 189–199.

Alvesson, M. and Sandberg, J., 2011. Generating Research Questions Through Problematization. *Academy of Management Review*, 36 (2), 247–271.

Applegate, L.M., 1999. Rigor and Relevance in Mis Research -- Introduction. *MIS Quarterly*, 23 (1), 1–2.

Archer, M.S., 2010. Routine, Reflexivity, and Realism. *Sociological Theory*, 28 (3), 272–303.

Archer, M.S., Bhaskar, R., Collier, A., Lawson, T., and Norrie, A., eds., 2007. *Critical realism: essential readings*. London; New York: Routledge.

Aristotle, 2000. *Nicomachean Ethics*. Cambridge, UK: Cambridge University Press.

Ashby, W.R., 1956. *An Introduction to Cybernetics*. Internet (1999). London: Chapman & Hall.

Bainbridge, L., 1983. Ironies of automation. *Automatica*, 19 (6), 775–779.

Balfe, N., Wilson, J.R., Sharples, S., and Clarke, T., 2012. Development of design principles for automated systems in transport control. *Ergonomics*, 55 (1), 37–54.

Baudrillard, J. and Glaser, S.F., 1994. *Simulacra and simulation*. Ann Arbor: University of Michigan Press.

BEA, Bureau d'Enquètes et d'Analyses pour la sécurité de l'aviation civile, 2012. *Final Report, On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France, flight AF 447 Rio de Janeiro - Paris*.

Beer, S., 1984. The Viable System Model: Its Provenance, Development, Methodology and Pathology. *The Journal of the Operational Research Society*, 35 (1), 7–25.

Bekier, M., Molesworth, B.R.C., and Williamson, A., 2011. Defining the drivers for accepting decision making automation in air traffic management. *Ergonomics*, 54 (4), 347–356.

Benbasat, I. and Zmud, R.W., 1999. Empirical Research in Information Systems: The Practice of Relevance. *MIS Quarterly*, 23 (1), 3–16.

Bertalanffy, L. von, 1950. An Outline of General Systems Theory. *The British Journal for the Philosophy of Science*, 1 (2), 134–165.

Bhaskar, R., 2008. *A realist theory of science*. New ed. London: Verso.

Billings, C.E., 1997. *Aviation automation: the search for a human-centered approach*. Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Birnboim, S., 2003a. The Automatic and Controlled Information-Processing Dissociation: Is It Still Relevant? *Neuropsychology Review*, 13 (1), 19–31.
Birnboim, S., 2003b. The Automatic and Controlled Information-Processing Dissociation: Is It Still Relevant? *Neuropsychology Review*, 13 (1), 19–31.
Boulding, K.E., 1956. General Systems Theory - The Skeleton of Science. *Management Science*, 2 (3), 197–208.
Box, G.E.P., 1976. Science and Statistics. *Journal of the American Statistical Association*, 71 (356), 791–799.
Brehmer, B., 1992. Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81 (3), 211–241.
Brehmer, B., 1994. Psychological aspects of traffic safety. *European Journal of Operational Research*, 75 (3), 540–552.
De Bruijn, H. and Herder, P.M., 2009. System and Actor Perspectives on Sociotechnical Systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 39 (5), 981 –992.
Cannon-Bowers, J.A., Salas, E., and Pruitt, J.S., 1996. Establishing the Boundaries of a Paradigm for Decision-Making Research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38 (2), 193–205.
Checkland, P., 1983. O.R. and the Systems Movement: Mappings and Conflicts. *Journal of the Operational Reasearch Society*, 34 (8), 661–675.
Checkland, P., 1985. From Optimizing to Learning: A Development of Systems Thinking for the 1990s. *The Journal of the Operational Research Society*, 36 (9), 757–767.
Checkland, P. and Scholes, J., 1999. *Soft Systems Methodology in Action*. Chichester: John Wiley & Sons, Ltd.
Chen, J.Y.C. and Terrence, P.I., 2009. Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics*, 52 (8), 907–920.
Cherns, A., 1976. The Principles of Sociotechnical Design. *Human Relations*, 29 (8), 783–792.
Cherns, A., 1987. Principles of Sociotechnical Design Revisted. *Human Relations*, 40 (3), 153–161.
Churchman, C.W., 1967a. Guest Editorial: Wicked Problems. *Management Science*, 14 (4), B141–142.
Churchman, C.W., 1967b. Wicked Problems. *Management Science*, Dec, p. B–141.
Churchman, C.W., 1970. Operations Research as a Profession. *Management Science*, 17 (2), B37–B53.
Clegg, C.W., 2000. Sociotechnical principles for system design. *Applied Ergonomics*, 31 (5), 463–477.
Collier, A., 1994. *Critical Realism, An Introduction to Roy Bhaskar's Philosophy*. London: Verso.
Conant, R., C. and Ashby, W.R., 1970. Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1 (2), 89–97.
Corley, K.G. and Gioia, D.A., 2011. Building Theory About Theory Building: What Constitutes a Theoretical Contribution? *Academy of Management Review*, 36 (1), 12–32.

Davenport, T.H. and Markus, M.L., 1999. Rigor Vs. Relevance Revisited: Response to Benbasat and Zmud. *MIS Quarterly*, 23 (1), 19–23.

Davis, M.S., 1971. That's Interesting! Towards a Phenomenology of Sociology and a Sociology of Phenomenology. *Philosophy of the Social Sciences*, 1 (2), 309–344.

Davis, M.S., 1986. 'That's Classic!' The Phenomenology and Rhetoric of Successful Social Theories. *Philosophy of the Social Sciences*, 16 (3), 285–301.

Dekker, S. and Hollnagel, E., 2004. Human factors and folk models. *Cognition, Technology & Work*, 6 (2), 79–86.

Dekker, S.W.A., 2007. Doctors Are More Dangerous Than Gun Owners: A Rejoinder to Error Counting. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49 (2), 177–184.

Dekker, S.W.A., 2013. On the epistemology and ethics of communicating a Cartesian consciousness. *Safety Science*, 56, 96–99.

Dekker, S.W.A. and Woods, D.D., 2002. MABA-MABA or Abracadabra? Progress on Human–Automation Co-ordination. *Cognition, Technology & Work*, 4 (4), 240–244.

Dobson, P.J., 2001. The Philosophy of Critical Realism - An Opportunity for Information Systems Research. *Information Systems Frontiers*, 3 (2), 199–210.

Dourish, P., 2004. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8 (1), 19–30.

Dreyfus, H. and S., 1986. Why Expert Systems Do Not Exhibit Expertise. *IEEE Expert*, 1 (2), 86 –90.

Dreyfus, H.L., 1972. *What computers can't do;: A critique of artificial reason,*. 1st ed. New York: Harper & Row.

Dreyfus, H.L., 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. 1st ed. The MIT Press.

Dreyfus, H.L. and Dreyfus, S.E., 1988. *Mind Over Machine*. Reprint. New York: Free Press.

Dreyfus, S.E. and Dreyfus, H.L., 1980. *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition*. University of California, Berkely: Operations Research Center, Research Report No. ORC-80-2.

Emmeche, C., Køppe, S., and Stjernfelt, F., 1997. Explaining Emergence: Towards an Ontology of Levels. *Journal for General Philosophy of Science / Zeitschrift für allgemeine Wissenschaftstheorie*, 28 (1), 83–119.

Endsley, M.R., 1995a. Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 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. and 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. and Kiris, E.O., 1995. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37 (2), 381–394.

Erlandsson, M. and Jansson, A., 2007. Collegial verbalization - a case study on a new method on information acquisition. *Behaviour & Information Technology*, 26 (6), 535–543.

Erlandsson, M. and Jansson, A., 2013. Verbal reports and domain-specific knowledge: a comparison between collegial and retrospective verbalisation. *Cognition, Technology & Work*, 15 (3), 239–254.

Faulkner, P. and Runde, J., 2013. Technological Objects, Social Positions, and the Transformational Model of Social Activity. *MIS Quarterly*, 37 (3), 803–818.

Flach, J.M., 1995. Situation Awareness: Proceed with Caution. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37 (1), 149–157.

Flood, R.L., 1990. Liberating Systems Theory: Toward Critical Systems Thinking. *Human Relations*, 43 (1), 49–75.

Flottau, J. and Wall, R., 2011. Painful Lessons. *Aviation Week & Space Technology*, 173 (20), 36.

Fogg, B.J., 2003. *Persuasive computing: technologies designed to change attitudes and behaviors*. San Francisco, Calif.; Oxford: Morgan Kaufmann ; Elsevier Science.

Forrester, J.W., 1971. Counterintuitive behavior of social systems. *Theory and Decision*, 2, 109–140.

Fox, W.M., 1995. Sociotechnical System Principles and Guidelines: Past and Present. *The Journal of Applied Behavioral Science*, 31 (1), 91–105.

Fracker, M.L., 1990. *Measures of Situation Awareness: Review and Future Directions*. Wright-Patterson AFB, OH: Human Engineering Division, Armstrong Laboratory, Final Report January 1990-January 1991 No. AL-TR-1991-0128.

Funke, G., Matthews, G., Warm, J.S., and Emo, A.K., 2007. Vehicle automation: A remedy for driver stress? *Ergonomics*, 50 (8), 1302–1323.

George, A.L. and Bennett, A., 2005. *Case Studies and Theory Development in the Social Sciences*. Fourth Printing edition. Cambridge, Mass: The MIT Press.

Gherardi, S., Turner, B.A., Pidgeon, N.E., and Ratzan, S.C., 1999. Man-Made Disasters 20 years later: Critical commentary. *Health, Risk & Society*, 1 (2), 233–239.

Giddens, A., 1984. *The Constitution of Society*. Cambridge: Polity Press.

Glaser, B.G., 2002. Conceptualization: On Theory and Theorizing Using Grounded Theory. *International Journal of Qualitative Methods*, 1 (2), 23–38.

Global Strategic Trends out to 2045 - Publications - GOV.UK [online], 2014. Available from: https://www.gov.uk/government/publications/global-strategic-trends-out-to-2045 [Accessed 28 Sep 2014].

Goldstein, B.E., 2007. *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*. New York: Cengage Learning.

Goldstein, J., 1999. Emergence as a Construct: History and Issues. *Emergence*, 1 (1), 49–72.

Graziano, A.M. and Raulin, M.L., 2007. *Research Methods, A Process of Inquiry*. 6th ed. Boston: Pearson Education, Inc.

Gregor, S., 2006. The Nature of Theory in Information Systems. *MIS Quarterly*, 30 (3), 611–642.

Grudin, J., 1990. The Computer Reaches Out: The Historical Continuity of Interface Design. *In*: *Proc. CHI 1990*. Presented at the CHI 1990, Seattle, WA, 261–268.

Gunetti, P., Thompson, H., and Dodd, T., 2011. Autonomous mission management for UAVs using soar intelligent agents. *International Journal of Systems Science*, 1–22.

Hancock, P.A., 2014. Automation: how much is too much? *Ergonomics*, 57 (3), 449–454.

Harrison, S., Tatar, D., and Sengers, P., 2007. The Three Paradigms of HCI. *In*: *alt.chi*.

Hjørland, B., 2009. Concept Theory. *Journal of the American Society for Information Science and Technology*, 60 (8), 1519–1536.

Hoc, J.-M., 2008. Cognitive ergonomics: a multidisciplinary venture. *Ergonomics*, 51 (1), 71–75.

Hollnagel, E. and Woods, D.D., 2005. *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton: Taylor & Francis Group.

Hollnagel, E., Woods, D.D., and Leveson, N., 2006. *Resilience Engineering, Concepts and Precepts*. Ashgate Publishing Limited.

Hoos, I.R., 1976. Engineers as Analysts of Social Systems: A Critical Enquiry. *Journal of Systems Engineering*, 4 (2), 81–88.

Hoyningen-Huene, P., 2006. CONTEXT OF DISCOVERY VERSUS CONTEXT OF JUSTIFICATION AND THOMAS KUHN. *In*: J. SCHICKORE and F. STEINLE, eds. *Revisiting Discovery and Justification*. Springer Netherlands, 119–131.

Huberman, A.M. and Miles, M.B., 1983. Drawing valid meaning from qualitative data: Some techniques of data reduction and display. *Quality and Quantity*, 17 (4), 281–339.

Huppertz, G., 2011. *Technology Forecast (TF) for Sub-System Type C02.08, Unmanned Land/Sea/Air Vehicles, Nano Air Vehicles*. Stockholm: Fraunhofer INT, commissioned by the Swedish Defence Material Administration, FMV, UNCLASSIFIED No. 11FMV2150-27.

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

Huxham, M. and Sumner, D., 2000. *Science and environmental decision making*. Harlow; New York: Prentice Hall.

INCOSE, 2006. *Systems Engineering Handbook, version 3*. INCOSE, No. INCOSE-TP-2003-002-03.

Jackson, M.C., 1991. The Origins and Nature of Critical Systems Thinking. *Systems Practice*, 4 (2), 131–149.

Jackson, M.C., 2000. *Systems approaches to management*. New York: Kluwer Academic / Plenum Publishers.

Jackson, M.C., 2001. Critical systems thinking and practice. *European Journal of Operational Research*, 128 (2), 233–244.

Jackson, M.C., 2010. Reflections on the development and contribution of critical systems thinking and practice. *Systems Research and Behavioral Science*, 27 (2), 133–139.

Jackson, M.C. and Keys, P., 1984. Towards a System of Systems Methodologies. *The Journal of the Operational Research Society*, 35 (6), 473–486.

Jansson, A., Olsson, E., and Erlandsson, M., 2006. Bridging the gap between analysis and design: Improving existing driver interfaces with tools from the framework of cognitive work analysis. *Cognition, Technology & Work*, 8, 41–49.

Jansson, A., Stensson, P., Bodin, I., Axelsson, A., and Tschirner, S., 2014. Authority and Level of Automation: Lessons to be learned in design of in-vehicle assistance systems. *In*: M. Kurosu, ed. *Human-Computer Interaction. Applications and Services*. Springer International Publishing, 413–424.

Jenkins, G.M., 1969. The Systems Approach. *Journal of Systems Engineering*, 1, 3–49.

Kaber, D.B. and Endsley, M.R., 1997. Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16 (3), 126–131.

Kahneman, D., 2003. A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58 (9), 697–720.

Kahneman, D., 2011. *Thinking, Fast and Slow*. Reprint. New York: Farrar, Straus and Giroux.

Kahneman, D. and Klein, G., 2009. Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64 (6), 515–526.

Kant, I., 1785. Fundamental Principles of the Metaphysic of Morals [e-text] [online]. Available from: http://www.gutenberg.org/etext/5682 [Accessed 6 May 2013].

Kavathatzopoulos, I., 2010. Robots and systems as autonomous ethical agents. *In*: *Towards creative technology for the 21st century*. Presented at the 11th International Conference on Intelligent Technologies, InTech 20108, Bangkok: Assumption University, 5–9.

Kirschenbaum, S.S., 2002. Uncertainty and Automation. *In*: *The Role of Humans in Intelligent and Automated Systems*. Presented at the NATO-RTO HFM Symposium, Warsaw, Poland: NATO RTO, 19–1–19–8.

Kirwan, B., 2000. Soft systems, hard lessons. *Applied Ergonomics*, 31 (6), 663–678.

Klein, G., 2008. Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50 (3), 456–460.

Klein, G.A., Orasanu, J., Calderwood, R., and Zsambok, C.E., 1993. *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Klein, G., Calderwood, R., and Clinton-Cirocco, A., 2010. Rapid Decision Making on the Fire Ground: The Original Study Plus a Postscript. *Journal of Cognitive Engineering and Decision Making*, 4 (3), 186–209.

Kruger, J. and Dunning, D., 1999. Unskilled and Unaware of it: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77 (6), 1121–1134.

Krupenia, S., Selmarker, A., Fagerlonn, J., Delsing, K., Jansson, A., Sandblad, B., and Grane, C., 2014. The 'Methods for Designing Future Autonomous Systems' (MODAS) project: Developing the cab for highly autonomous truck. *In*: *Advances in Human Factors and Ergonomics 2014. 20 Volume*

Set: Proceedings of the 5th AHFE Conference 19-23 July 2014. Presented at the AHFE2014, Krakow, 70–81.
KSU, 2006. Forsmarksincidenten den 25 Juli 2006. *Bakgrund*, 19 (5).
KSU, 2007. The Forsmark incident 25th July 2006. *Bakgrund*, 20 (1).
Laaksoharju, M., 2010. Let Us Be Philosophers! Computerized Support for Ethical Decision Making. IT Licentiate thesis 2010-005. Department of Information Technology, Uppsala University.
Laaksoharju, M., 2014. Designing for Autonomy. Doctoral dissertation. Uppsala.
Lanier, J., 2011. *You are not a gadget*. Paperback Edition. London: Penguin Books Ltd.
Leveson, N., 2004a. A Systems-Theoretic Approach to Safety in Software-Intensive Systems. *IEEE Transactions on Dependable and Secure Computing*, 1 (1), 66–86.
Leveson, N., 2004b. A new accident model for engineering safer systems. *Safety Science*, 42 (4), 237–270.
Leveson, N., Dulac, N., Marais, K., and Carrol, J., 2009. Moving Beyond Normal Accidents and High Reliability Organizations; A Systems Approach to Safety in Complex Systems. *Organization Studies*, 30 (02&03), 227–249.
Leveson, N.G., 1986. Software safety: why, what, and how. *ACM Comput. Surv.*, 18 (2), 125–163.
Lieberman, H.R., Bathalon, G.P., Falco, C.M., Morgan, C.A., Niro, P.J., and Tharion, W.J., 2005. The Fog of War: Decrements in Cognitive Performance and Mood Associated with Combat-Like Stress. *Aviation, Space, and Environmental Medicine*, 76 (7), C7–C14.
Lilienfeld, R., 1975. Systems Theory as an Ideology. *Social Research*, 42, 637–660.
Lintern, G., 2005. What is a cognitive system? *In*: *Proceedings of the Fourteenths International Symposium on Aviation Psychology*. Presented at the ISAP 2005, Dayton, OH, 398–402.
Marchal, J.H., 1975. On the Concept of a System. *Philosophy of Science*, 42 (4), 448–468.
McKaughan, D.J., 2008. From Ugly Duckling to Swan: C. S. Peirce, Abduction, and the Pursuit of Scientific Theories. *Transactions of the Charles S. Peirce Society*, 44 (3), 446–468.
Miles, M.B. and Huberman, A.M., 1984. Drawing Valid Meaning from Qualitative Data: Toward a Shared Craft. *Educational Researcher*, 13 (5), 20–30.
Miles, M.B. and Huberman, A.M., 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd ed. SAGE Publications, Inc.
Miller, G.A., 1955. The Magical Number Seven, Plus or Minus Two, Some Limits on Our Capacity for Processing Information. *Psychological Review*, 101 (2), 343–352.
Mingers, J., 2000. The contribution of critical realism as an underpinning philosophy for OR/MS and systems. *Journal of the Operational Reasearch Society*, 51 (11), 1256–1270.
Mingers, J., 2001. Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, 12 (3), 240–259.
Mingers, J., 2003. The paucity of multimethod research: a review of the information systems literature. *Information Systems Journal*, 13 (3), 233–249.

Mingers, J., 2004. Paradigm wars: ceasefire announced who will set up the new administration? *Journal of Information Technology*, 2004 (19), 165–171.

Mingers, J., Mutch, A., and Willcocks, L., 2013. Critical Realism in Information Systems Research. *MIS Quarterly*, 37 (3), 795–802.

Morris, C.W., 1980. Human Autonomy and the Natural Right to be Free. *The Journal of Libertarian Studies*, IV (4), 379–392.

Müller, M., 2011. *Technology Forecast (TF) for Sub-System Type C02.08, Unmanned Land/Sea/Air Vehicles, Biomimetic UUV*. Stockholm: Fraunhofer INT, commissioned by the Swedish Defence Material Administration, FMV, UNCLASSIFIED No. 11FMV2150-23.

Myers, M.D. and Klein, H.K., 2011. A Set of Principles for Conducting Critical Research in Information Systems. *MIS Quarterly*, 35 (1), 17–36.

NAIIC, 2012. *The official report of The Fukushima Nuclear Accident Independent Investigation Commission*. The National Diet of Japan Fukushima Nuclear Accident Independent Investigation Commission.

National Transportation Safety Board, 2010. *Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River, US Airways Flight 1549, Airbus A320-214, N106US, Weehawken, New Jersey, January 15, 2009*. Washington DC, No. Aircraft Accident Report NTSB/AAR-10 /03.

Nonaka, I., Toyama, R., and Konno, N., 2000. SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. *Long Range Planning*, 33 (1), 5–34.

Norman, D.A., 2002. *The design of everyday things*. New York: Basic Books.

Olson, G.M. and Olson, J.S., 2000. Distance Matters. *Human-Computer Interaction*, 15 (2/3), 139–178.

Orlikowski, W.J., 1992. The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science*, 3 (3), 398–427.

Orlikowski, W.J. and Iacono, C.S., 2001. Research Commentary: Desperately Seeking the 'IT' in IT Research - A Call to Theorizing the IT Artifact. *Information Systems Research*, 12 (2), 121–134.

Parasuraman, R., 2000. Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, 43 (7), 931–951.

Parasuraman, R., Sheridan, T.B., and Wickens, C.D., 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 30 (3), 286 – 297.

Parasuraman, R., Sheridan, T.B., and 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.

Parker, D., Lawrie, M., and Hudson, P., 2006. A framework for understanding the development of organisational safety culture. *Safety Science*, 44, 551–562.

Parker, D., Manstead, A.S.R., Stradling, S.G., and Reason, J.T., 1992. Intention to Commit Driving Violations: An Application of the Theory of Planned Behavior. *Journal of Applied Psychology*, 77 (1), 94–101.

Pava, C., 1986. Redesigning Sociotechnical Systems Design: Concepts and Methods for the 1990s. *The Journal of Applied Behavioral Science*, 22 (3), 201–221.
Perrow, C., 1999. *Normal Accidents: living with high-risk technologies*. New Jersey: Princeton University Press.
Perrow, C., 2004. A Personal Note on Normal Accidents. *Organization & Environment*, 17 (1), 9–14.
Polanyi, M., 1966. *The Tacit Dimension*. 2009th ed. Chicago: The University of Chicago Press.
Rachels, J., 2009. *The Elements of Moral Philosophy*. London: McGraw-Hill Higher Education.
Reason, J., 1990. *Human Error*. Cambridge University Press.
Reason, J., 1998. Achieving a safe culture: theory and practice. *Work & Stress*, 12 (3), 293–306.
Reason, J., 2000. Safety paradoxes and safety culture. *Injury Control & Safety Promotion*, 7 (1), 3–14.
Reason, J., 2008. *The Human Contribution*. Ashgate Publishing Limited.
Reed, E.S., 1996. *The Necessity of Experience*. New Haven and London: Yale University Press.
Rittel, H.W. and Webber, M.M., 1973. Dilemmas in a General Theory of Planning. *Policy Sciences*, 4, 155–169.
Röttger, S., Bali, K., and Manzey, D., 2009. Impact of automated decision aids on performance, operator behaviour and workload in a simulated supervisory control task. *Ergonomics*, 52 (5), 512–523.
Ruhlig, K., 2011. *Technology Forecast (TF) for Sub-System Type C02.08, Unmanned Land/Sea/Air Vehicles, Biomimetic UUV*. Stockholm: Fraunhofer INT, commissioned by the Swedish Defence Material Administration, FMV, UNCLASSIFIED No. 11FMV2150-24.
Ryle, G., 1945. Knowing How and Knowing That: The Presidential Address. *Proceedings of the Aristotelian Society*, 46, 1–16.
Sagan, S.D., 2004a. Learning from Normal Accidents. *Organization & Environment*, 17 (1), 15–19.
Sagan, S.D., 2004b. The Problem of Redundancy Problem: Why More Nuclear Security Forces May Produce Less Nuclear Security. *Risk Analysis: An International Journal*, 24 (4), 935–946.
Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., Jenkins, D.P., McMaster, R., and 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.
Sandberg, J. and Alvesson, M., 2011. Ways of constructing research questions: gap-spotting or problematization? *Organization*, 18 (1), 23–44.
Sarter, N.B. and 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. and 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), 5–19.

Sauer, J., Kao, C.-S., and Wastell, D., 2012. A comparison of adaptive and adaptable automation under different levels of environmental stress. *Ergonomics*, 55 (8), 840–853.

Sauer, J., Kao, C.-S., Wastell, D., and Nickel, P., 2011. Explicit control of adaptive automation under different levels of environmental stress. *Ergonomics*, 54 (8), 755–766.

Schön, D.A., 1983. *The reflective practitioner: how professionals think in action*. New York: Basic Books.

Searle, J.R., 1980. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3 (03), 417–424.

Sheridan, T.B., 2008. Risk, Human Error, and System Resilience: Fundamental Ideas. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50 (3), 418–426.

Shiffrin, R.M. and Schneider, W., 1977a. Controlled and Automatic Human Information Processing: I. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review*, 84 (1), 1–66.

Shiffrin, R.M. and Schneider, W., 1977b. Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review*, 84 (2), 127–190.

SHK, Swedish Board of Accident Investigations, 1993. *Luftfartshändelse den 27 december 1991 i Gottröra, AB län*. No. C 1993:57.

Simon, H.A., 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106 (6), 467–482.

Smith, M.L., 2006. Overcoming theory-practice inconsistencies: Critical realism and information systems research. *Information and Organization*, 16 (3), 191–211.

Squire, P.N. and Parasuraman, R., 2010. Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment. *Ergonomics*, 53 (8), 951–961.

Stanovich, K.E. and West, R.F., 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23 (05), 645–665.

Stanton, N.A., Salmon, P.M., Walker, G.H., and Jenkins, D.P., 2010. Is situation awareness all in the mind? *Theoretical Issues in Ergonomics Science*, 11 (1-2), 29–40.

Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M.S., Linsell, M., Dymott, R., and 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. and Young, M.S., 1998. Vehicle automation and driving performance. *Ergonomics*, 41 (7), 1014–1028.

Stanton, N.A. and Young, M.S., 2005. Driver behaviour with adaptive cruise control. *Ergonomics*, 48 (10), 1294–1313.

Steele, G.R., 2005. Critical thoughts about critical realism. *Critical Review: A Journal of Politics and Society*, 17 (1-2), 133–154.

Stensson, P., 2010. Thoughts about the Consequences of Inappropriate Application of Systems Engineering. *In*: *Proceedings of the 7th European Systems Engineering Conference*. Presented at the EuSEC 2010, Stockholm: INCOSE.

Stensson, P. and Jansson, A., 2014a. Autonomous technology – sources of confusion: a model for explanation and prediction of conceptual shifts. *Ergonomics*, 57 (3), 455–470.

Stensson, P. and Jansson, A., 2014b. Edge Awareness - A Dynamic Safety Perspective on Four Accidents/Incidents. *In*: B. Amaba and B. Dalgetty, eds. *Advances in Human Factors, Software and Systems Engineering*. Krakow: AHFE Conference, 168–179.

Sterman, J.D., 2002. All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review*, 18 (4), 501–531.

Suchman, L.A., 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge, UK: Cambridge University Press.

Sullins, J.P., 2002. The Ambiguous Ethical Status of Autonomous Robots [online]. Available from: https://www.aaai.org/Papers/Symposia/Fall/2005/FS-05-06/FS05-06-019.pdf [Accessed 23 May 2009].

Sutton, R.I. and Staw, B.M., 1995. What Theory is Not. *Administrative Science Quarterly*, 40 (3), 371–384.

Taleb, N.N., 2010. *The Black Swan, The Impact of the Highly Improbable*. Second Edition. New York: Random House.

Technology Forecast 2012 Military utility of ten technologies - a report from seminars at the SNDC Department of Military-Technology, 2012. Stockholm: FMV, Swedish Defence Material Administration, UNCLASSIFIED No. 366185-LB835880.

Thomas, G., 2011. Wake-Up Call, The lessons of AF447 and other recent high-automation aircraft incidents have wide training implications. *Air Transport World*, 39–43.

Trist, E., 1981. The evolution of socio-technical systems, a conceptual framework and an action research program. *Journal of the Royal College of General Practitioners*, 2.

Trist, E.L. and Bamforth, K.W., 1951. Some Social and Psychological Consequences of the Longwall Method of Coal-Getting An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System. *Human Relations*, 4 (1), 3–38.

Turing, A.M., 1950. Computing Machinery and Intelligence. *Mind*, 59 (236), 443–460.

Turner, B.A. and Pidgeon, N.F., 1997. *Man-made disasters*. Oxford: Butterworth-Heinemann.

Tversky, A. and Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185 (4157), 1124–1131.

Tversky, A. and Kahneman, D., 1981. The Framing of Decisions and the Psychology of Choice. *Science*, 211, 453–458.

Ulrich, W., 1987. Critical heuristics of social systems design. *European Journal of Operational Research*, 31 (3), 276–283.

Ulrich, W., 2003. Beyond methodology choice: critical systems thinking as critically systemic discourse. *Journal of the Operational Reasearch Society*, 54, 325–342.

Ulrich, W., 2007. Viewpoints: Philosophy for professionals: towards critical pragmatism. *Journal of the Operational Reasearch Society*, 58, 1109–1113.

Vicente, K.J., 1990. Coherence- and correspondence-driven work domains: implications for systems design. *Behaviour & Information Technology*, 9, 493–502.

Vicente, K.J., 1999. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Vicente, K.J., 2002. Ecological Interface Design: Progress and Challenges. *Human Factors*, 44 (1), 62–78.

Vicente, K.J., 2006. *The Human Factor*. New York: Routledge.

Vicente, K.J., 2011. *Human-tech: ethical and scientific foundations*. New York: Oxford University Press.

Vicente, K.J. and Rasmussen, J., 1992. Ecological Interface Design: Theoretical Foundations. *IEEE Transactions on Systems, Man, and Cybernetics*, 22 (4), 589–606.

Walker, G.H., Stanton, N.A., Salmon, P.M., and Jenkins, D.P., 2008. A review of sociotechnical systems theory: a classic concept for new command and control paradigms. *Theoretical Issues in Ergonomics Science*, 9 (6), 479–499.

Walsham, G., 1995. Interpretive case studies in IS: nature and method. *European Journal of Information Systems*, 4 (2), 74–81.

Waraich, Q.R. ('Raza'), Mazzuchi, T.A., Sarkani, S., and Rico, D.F., 2013. Minimizing Human Factors Mishaps in Unmanned Aircraft Systems. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 21 (1), 25–32.

Weaver, W., 1948. Science and complexity. *American Scientist*, 36.

Weber, R., 2003. The Problem of the Problem. *MIS Quarterly*, 27 (1), 1–1.

Weber, R., 2012. Evaluating and Developing Theories in the Information Systems Discipline. *Journal of the Association for Information Systems*, 13 (1), 1–30.

Weick, K.E., 1989. Theory Construction as Disciplined Imagination. *Academy of Management Review*, 14 (4), 516–531.

Weick, K.E., 1995. What Theory Is Not, Theorizing Is. *Administrative Science Quarterly*, 40 (3), 385–390.

Weick, K.E., 2004. Normal Accident Theory As Frame, Link, and Provocation. *Organization & Environment*, 17 (1), 27–31.

Weick, K.E., Sutcliffe, K.M., and Obstfeld, D., 2005. Organizing and the Process of Sensemaking. *Organization Science*, 16 (4), 409–421.

Weizenbaum, J., 1976. *Computer Power and Human Reason*. W. H. Freeman and Company.

Westrum, R., 2004a. A typology of organisational cultures. *Quality and Safety in Health Care*, 13 (suppl_2), ii22–ii27.

Westrum, R., 2004b. A typology of organisational cultures. *Quality and Safety in Health Care*, (13), ii22–ii27.

Whetten, D.A., 1989. What Constitutes a Theoretical Contribution? *Academy of Management Review*, 14 (4), 490–495.

Wickens, C.D., 2008. Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50 (3), 397–403.

Wickens, C.D. and Hollands, J.G., 2000. *Engineering Psychology and Human Performance*. New Jersey: Prentice-Hall Inc.

Wiener, N., 1948. *Cybernetics or control and communication in the animal and the machine*. Cambridge, Mass.: The Technology Press.

Wilde, G.J.S., 1998. Risk homeostasis theory: an overview. *Injury Prevention*, 4 (2), 89–91.

Woods, D.D. and Hollnagel, E., 2006. *Joint Cognitive Systems: Patterns in Cognitive Systems*. Boca Raton, FL: Taylor & Francis.

Wynn, D.E.J. and Williams, C.K., 2008. Critical Realm-Based Explanatory Case Study in Information Systems. *In*: *ICIS 2008*. Presented at the Twenty Ninth International Conference on Information Systems, Paris.

Wynn, J., Donald and Williams, C.K., 2012. Principles for Conducting Critical Realist Case Study Research in Information Systems. *MIS Quarterly*, 36 (3), 787–810.

Yates, F.E., 1978. Complexity and the limits to knowledge. *The American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, (235), 201–204.

Yeung, H.W.C., 1997. Critical realism and realist research in human geography: a method or a philosophy in search of a method? *Progress in Human Geography*, 21 (1), 51–74.

Young, M.S. and Stanton, N.A., 2007a. What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics*, 50 (8), 1324–1339.

Young, M.S. and Stanton, N.A., 2007b. Back to the future: Brake reaction times for manual and automated vehicles. *Ergonomics*, 50 (1), 46–58.

Zachariadis, M., Scott, S., and Barrett, M., 2013. Methodological Implications of Critical Realism for Mixed-Methods Research. *MIS Quarterly*, 37 (3), 855–879.