Recovery uncovered : how people in the chemical process industry recover from failures

DOI:
10.6100/IR572823

Document status and date:
Published: 01/01/2004

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Recovery uncovered:

How people in the chemical process industry recover from failures

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Kanse, Lisette

Recovery uncovered: How people in the chemical process industry recover from failures / by Lisette Kanse. – Eindhoven: Technische Universiteit Eindhoven, 2004. – Proefschrift.
ISBN 90-386-1788-7
NUR 807
Keywords: Failure recovery / Human reliability / Safety management / Near misses / Incident analysis / Chemical industry

Printed by Eindhoven University Press

Recovery uncovered:
How people in the chemical process industry recover from failures

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven,
op gezag van de Rector Magnificus, prof.dr. R.A. van Santen, voor een commissie
aangewezen door het College voor Promoties in het openbaar te verdedigen
op woensdag 7 januari 2004 om 16.00 uur

door

Lisette Kanse

geboren te Terneuzen

Dit proefschrift is goedgekeurd door de promotoren:
prof.dr. C.G. Rutte
en
prof.dr. E. Hollnagel

Copromotor:
dr. T.W. van der Schaaf

Preface

As with many things in life, the idea of doing a PhD had to grow on me for a while, before I realized that this definitely was the way to go. On the road to making it happen, I have been helped by a lot of people. In the first place by my supervisor Tjerk van der Schaaf, the man with the many ideas and an amazing network, the person who found most of the jobs I have had as a grown-up for me. He knew the research I have done was worth doing long before I started my academic life. Also by my promotor Christel Rutte. If it wasn't for her, my PhD research might not have started at all. I admire her for taking in a stranger with a strange PhD research topic – at least that is what I was to her at the time. By my former employer, David Embrey, from Human Reliability Associates: thanks for letting me go and staying interested at the same time. By my second promotor, Erik Hollnagel, who managed to find time for me regardless of the large number PhD students he was already supervising, and who has always helped me to believe I was on the right track. Thanks for many enlightening meetings, including those with PhD students whose research resembled mine much more than that of most PhD students within my own group.

I also want to thank Thomas Bove, who completed his PhD thesis on error recovery in 2002 at Risø National Laboratory and the University of Roskilde, for the many inspirational discussions we have had in the recent years. It was great to be able to share insights with someone who felt equally passionate about recovery processes as a research topic.

No research without research data. I owe a debt of gratitude to the four chemical plants who participated in my studies and who shall remain nameless in this thesis. Many thanks to all who helped me during data collection: safety staff, training coordinators, managers, and last but not least operators, technicians and other personnel who talked to me about things that almost went really wrong but were recovered, or who filled out lists of questions about such situations. Trust me, I know these are not the easiest situations to think and talk about.

The good thing about doing a PhD within a research school, in a university, is the environment you get to work in. Over the years that I've been there, all my coworkers in the Human Performance Management group, knowingly or unknowingly, each have taught me something. I want to thank them all for the wise lessons. I specifically want to thank Ad Kleingeld, not only for his share in these lessons, but also for understanding what it is like to be an engineer in a group of mostly psychologists (but at least a 'homegrown' engineer, in our case). Extra thanks go to the T crew for their company and friendship and the platform they provided to let off steam every now and then; and to the other PhD students from our group who also believed lunch breaks should occasionally be spent on fitness. My roommate Heleen van Mierlo, who had to put up with me since day one (and I don't think

she was given a choice), who taught me a lot about work and organizational psychology, and who inspiringly demonstrated that IT CAN BE DONE, deserves very special thanks, for everything, also simply for being there. I will miss her very much.

Outside the group of my immediate coworkers but still within the same university, Lesley Smits, who worked on a design project for the department of Chemical Engineering and Chemistry in one of the plants that participated in my research, deserves many thanks for her willingness to include my research project within hers and for her assistance in data collection.

I owe special thanks to Harry Garst, Ad de Jong, and Lieuwe Dijkstra, who compensated for my initial lack of knowledge regarding statistical methods.

In addition, I want to thank everybody who played a role in reviewing earlier versions of this thesis for their insightful suggestions.

I have to thank the people outside my work environment for keeping me level-headed: my waterpolo club, friends, and of course my and my husbands family. Plus Bart, who is always in my heart even when he is not nearby. His belief in me got me here.

Contents

List of figures

List of tables

1 Introduction

The insight that "if anything can go wrong, it will", has both inspired and frustrated many of those who have been responsible for safety and reliability in work- and other environments, probably even since before this insight was named 'Murphy's law'. As the story goes (see http://www.murphys-laws.com/murphy/murphy-true.html), Captain Edward A. Murphy was an engineer at Edwards Air Force Base in California, USA, working on a project aimed at finding out how much sudden deceleration a person can stand in a crash. One day in 1949, a technician had wired a transducer wrongly, rendering an entire test useless. Captain Murphy found out and furiously said about the technician: "if there is any way to do it wrong, he'll find it". The project manager, who kept a list of 'laws' like this, added this one and named it Murphy's law, thus giving an old law that had been around for many years a name. At a press conference shortly afterwards, an Air Force doctor involved in the project proclaimed that the good safety record on the project was due to a firm belief in Murphy's law and the necessity to try and circumvent it. This was the first of many quotations of Murphy's law. Several versions of it have been put forward over the years, the one given at the start of this paragraph being the most frequently used. Whether this story tells us the true origins of Murphy's law or not is unclear, but this does not alter the fact that this name has become widely used to describe the phenomenon of things failing, often in unforeseen ways, at the worst possible time.

Regardless of their awareness of and views on Murphy's law, traditionally the focus of people working in safety and reliability management has mainly been on the prevention of both technical failures, human failures or errors, and more recently also organizational or management failures (Reason, 1991). However, additional benefits can be gained from focusing on what can be done after a failure has occurred, but before this leads to negative consequences (Frese, 1991; Kontogiannis, 1999). After all, not all failures, no matter which type, can be foreseen, and even foreseen failures can not always be prevented – the appropriate preventive measures may be unknown, impossible to implement, or the benefits of implementing them may not outweigh the costs. And as Murphy's law suggests, these failures will occur. What we really want to prevent is not so much the failures themselves, but rather the negative consequences of these failures. Even if Murphy's law is true, all is not lost.

Generally, a failure or combination of several different or similar types of failures initially only leads to an unwanted situation, a deviation from normal, from an intended or desired state. In most cases, both in work and everyday life, there is still a chance to recover from the failure(s) and the resulting deviation, thus returning the situation back to normal, before any negative consequences take place. The following examples may help to clarify the difference between a failure, the resulting deviation, and its potential negative consequences:

- a defective engine (failure), if unnoticed, can overheat (deviation), and lead to a fire and thus damage, production loss and potential injuries (negative consequences);
- forgetting to refill lubrication oil (failure) on rotating equipment can lead to unlubricated operation (deviation), which can go on for a while before eventually leading to a breakdown causing both production and product loss (negative consequences);
- if the report that spare equipment has been taken into service does not reach the maintenance department (failure), they will not repair the equipment that was replaced with the spare, and operation now continues without a back-up for this critical piece of equipment being readily available (deviation), which can lead to major production - and related safety problems (negative consequences) as soon as the currently active spare equipment also fails.

Figure 1 contains a graphical representation of this sequence from failures to a deviation to negative consequences. In this sequence, the deviation is how the problem situation resulting from the failures has *actually manifested* itself, and the negative consequences are the *potential effects* to which the situation may lead, even though it is not certain that these will occur. A more detailed discussion and further definitions of each of the elements in this sequence follow in the next chapter.

Figure 1 Relationship between failures, deviation, and negative consequences

People are often simply regarded as those elements in a system which commit errors (i.e. human failures) and thus make things go wrong; they are often the first to whom blame for any failure will be assigned, justified or not (e.g. Reason, 1990). However, in many cases people may also be the ones who detect that something is wrong and who take the appropriate countermeasures (e.g. Rouse, 1981), thereby preventing things from getting further out of hand. This means that via such recovery actions people may contribute to an organization's safety and reliability in a positive way, thus serving not only as a weak, but also as a strong link.

A recovery process, in my definition, includes both the detection that something is wrong (detection of the deviation), and all the information processing and actions performed in response to this to avoid negative consequences, until the situation is returned to normal, that is, the intended or desired state. A graphical representation of how recovery intervenes in the sequence of failures, deviation, and potential negative consequences is given in figure 2.

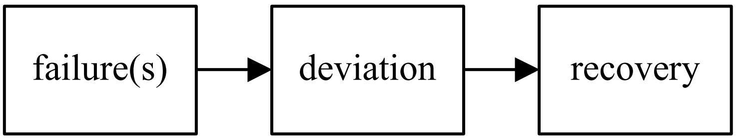


Figure 2 Recovery instead of allowing negative consequences to occur

Figures 1 and 2 also show the difference between accidents and near misses. Accidents are those events where a failure or combination of several different or similar types of failures has eventually led to at least some undesired safety-, reliability-, or otherwise performance-related negative consequences, like in figure 1. In events where all the necessary recovery actions have been performed in a timely and effective manner, such that undesired negative consequences have been avoided, like in figure 2, a near miss results (e.g. Van der Schaaf, Lucas & Hale, 1991). In fact, near misses are also possible in situations where no recovery actions have been performed, where the potential negative consequences that could have occurred were avoided by sheer luck or coincidence. In this thesis, incidents are defined as all events where failures have occurred with the potential to lead to negative safety-, reliability-, or otherwise performance-related consequences, irrespective of whether in the end these negative consequences became manifest, at least to some extent, or were avoided completely. This definition encompasses both accidents and near misses. Appendix 1 contains a glossary of these and other key terms that are introduced, defined and used throughout the rest of this thesis.

As will become obvious in the next chapter, for a number of decades already researchers in fields like human-machine interaction and human decision making have been aware of the role people can play in detecting, diagnosing, and correcting failures, ranging from human errors to technical or system faults. The next chapter will also show, however, that detailed insights in the steps involved in the processes people in actual work situations follow to recover from different types and combinations of failures, the factors on which these steps and complete processes depend, and how this all relates to process outcomes, are still limited. This gap in available knowledge is unfortunate, as such insights can be used to promote recovery possibities and can thus help to improve organizational safety and reliability and even performance in general. The topic of this PhD thesis was chosen based on this consideration, combined with the understanding that people may play an important role in recovery processes. The overall question that has guided the research described in the rest of this thesis was:

> *Which processes do people follow to recover from different types and combinations of failures in work situations, and what influences these processes and their outcomes?*

To further specify this relatively global overall research question, a literature review was performed to identify the relevant existing insights into recovery processes and areas in need of further research. The next chapter contains an overview of these existing insights and gaps in knowledge. Based on this overview, throughout that chapter, seven more specific research questions are introduced. The first five specific research questions cover several specific aspects of the recovery process itself; they aim at establishing what is done, why, how, when, for how long, by whom, and why by those persons, during a recovery process. The sixth specific research question aims at explaining what happens during a recovery process (cf. the first five research questions) by establishing how the preceding types and combination of failures, the potential consequences of the deviation to which these failures led, and the moment of detection influence the recovery processes that follows. The seventh and last specific research question aims at establishing the recovery process end results and how these can be explained by both the events leading up to the recovery process (cf. research question six) and the recovery process itself (cf. the first five research questions).

Three initial empirical studies were performed in two chemical plants to answer the seven specific research questions presented in chapter 2. The rationale and research design for these studies are presented in the third chapter, followed by findings, discussion and conclusions based on these studies, for each of the seven specific research questions. In this chapter, readers who are mainly interested in the conclusions from the studies can safely skip the sections where the findings are presented. The fourth chapter describes the rationale, research design, and findings complete with discussion and conclusions, for a more in-depth exploration of two of the specific research questions, for which two additional empirical studies were performed in two additional chemical plants. The findings of all five studies are summarized and their implications reviewed in the fifth and final chapter.

2 Literature review[1]

2.1 Purpose and outline

The purpose of this chapter is to set the scene for the research described in this thesis. Drawing upon existing literature mainly from the domains of work- and organizational psychology, cognitive psychology, human reliability, ergonomics, safety management, problem solving and (naturalistic) decision making, in this chapter I present what I consider to be the relevant current insights into recovery processes and the factors influencing these processes and their outcomes. The more detailed, specific research questions that follow from this literature review and that have guided the research for this thesis are also presented throughout the next sections.

As recovery processes are the main research theme in this thesis, this chapter starts with a review of what exactly happens during these processes, by looking for answers to questions such as: what is done during a recovery process, why, how, when, for how long, by whom, and why by those persons. Both what is currently known and what needs further research with regard to these questions is addressed in section 2.2. Within that section, subsection 2.2.1 deals with the questions what is done during a recovery process, why, and when, by addressing the different possible steps involved in the process, their goals, and the order in which they are performed. Subsection 2.2.2 deals with the question how, by addressing the extent to which these recovery process steps involve ad-hoc or pre-planned behavior, and the amount of conscious reasoning involved. Subsection 2.2.3 addresses the question how long each of the process steps and the entire process take. Subsection 2.2.4 deals with prioritization of recovery steps, covering another aspect of the questions when and how long. Subsection 2.2.5 deals with the people involved in the process and addresses the questions by whom the recovery steps are performed and why by those persons.

As part of the explanation of what happens during a recovery process may lie outside the time frame during which the recovery steps are performed, in section 2.3 the scope of the research is widened to include the events that precede a recovery process, that have led to a situation from which recovery is needed and thus may have influenced the recovery process. The preceding types and combination of failures, the potential consequences of the deviation to which these failures led, and the moment of detection are addressed, along with their impact on the recovery process that follows.

---

[1] Parts of this review have been presented in: Van der Schaaf & Kanse (1999; 2000); Kanse & Van der Schaaf (2000a; 2000b; 2001b; 2001c); Kanse (2002); Kanse, Van der Schaaf, & Rutte (2003).

Finally, in section 2.4, the focus shifts to what is known with regard to the various possible outcomes or end results of recovery processes, thus extending the scope of the research to also include the situation after the recovery process is completed. These process outcomes may depend both on characteristics of the events leading up to a recovery process and on characteristics of the recovery process itself.

Slightly extending the sequence presented in figure 2, figure 3 demonstrates how the topics mentioned above are linked together and indicates in which sections each of these topics are addressed in more detail.

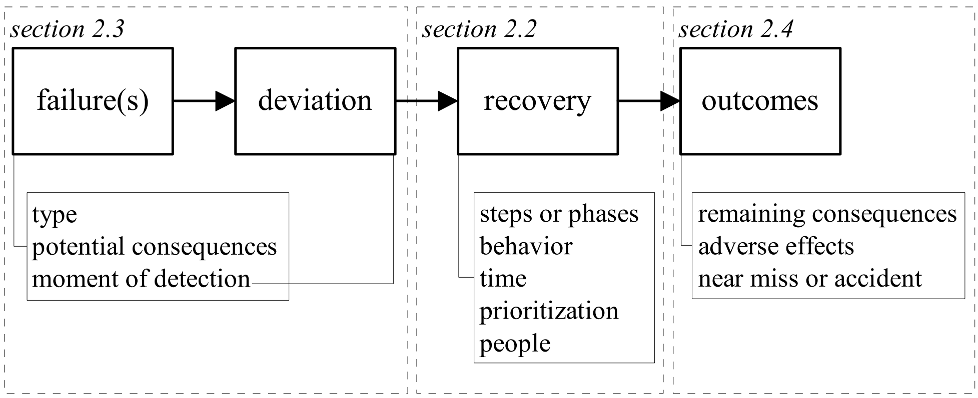
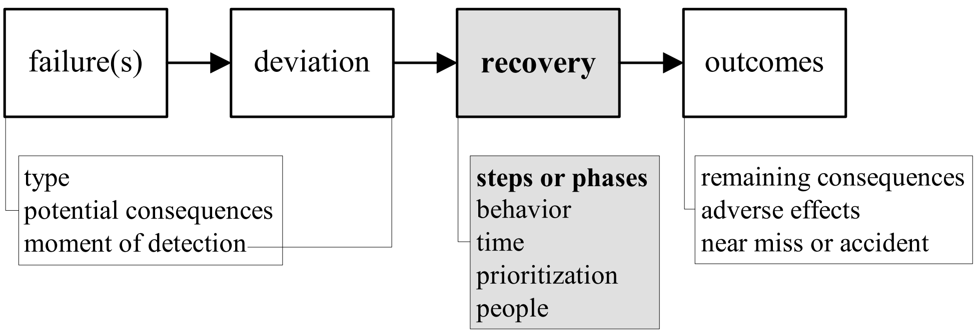


Figure 3 Antecedents, recovery processes, and outcomes; and sections in which these topics are addressed

2.2 The recovery process

2.2.1 Steps or phases involved in recovery



One of the insights stemming from research on human motivation is the theory on goal hierarchies (e.g. Geen, 1995), which explains how people can reach a higher-level goal or desired state via attainment of the lower level goals associated with that desired state. In other words, to reach an overall, higher-level goal, people generally have to perform a sequence of steps, each with their own specific subgoal, via which part of the overall goal is achieved. In the case of recovery processes, the recognition that an ongoing higher-level goal is not being or about to be reached has triggered a new higher-level, overall goal: that of recovery, in other words to return to the normal, intended situation while avoiding negative consequences. In line with the theory on goal hierarchies, one would expect that any recovery process consists of a number of separate steps or phases with separate subgoals, that have to be completed in order to return to the normal, intended situation and thus attain the overall goal.

In his model of stages and levels in human-machine interaction, Norman (1984) also referred to the existence of different levels of specification of goals, ranging from high-level to very detailed, specific goals. He argued that each of the stages of activity he distinguished, that is, intention, selection, execution, and evaluation (in more recent versions even more stages are distinguished), may occur at various levels of specification.

The notion of people going through a sequence of phases can also be found in problem solving and decision making theory and models. In their book on human problem solving, for example, Newell and Simon (1972) argued that a problem solving process consists of a sequence of episodes or distinct segments of behavior associated with attaining a goal. For goals that are not too complicated, a task analysis can reveal which episodes are necessary to attain that goal. Each of these episodes has its own subgoal. The result of pursuing one subgoal may even be used as input for an episode that follows later.

Recovering from failures can be seen as a special kind of problem solving, namely solving the problems that are the immediate result of a failure or combination of several different or similar types of failures. Kepner and Tregoe (1981) defined problems as deviations from expected performance, of people, systems, policies or equipment. As was pointed out in the first chapter, the occurrence of a failure or combination of several different or similar types of failures can lead to such deviations. If that is the case, then the problem solving process and the recovery process are exactly the same thing.

The phase theorem which dominates much of the literature on problem solving (for a review see Lipshitz & Bar-Ilan, 1996) does not only suggest that people follow several phases in the process of solving problems, but also that certain sequences of phases are more likely to be successful than others. Over the years, a variety of problem solving or decision making phase models have been proposed. Table 1 contains a few examples.

Table 1 Some examples of phase models from the domain of problem solving and decision making

Simon (1960)	**Newell & Simon (1972)**	**Mintzberg, Raisingahni & Theoret (1976)**	**Kepner & Tregoe (1981)**	**Carroll & Johnson (1990)**	**Lipshitz & Bar-Ilan (1996)**
intelligence, aiming at understanding the problem; ***design*** of alternative solutions; ***choice*** of the best solution.	***selecting a goal***, by producing an internal representation of the problem; ***selecting a method*** for (a step towards) solving the problem; ***applying the method***; ***evaluating the results*** of the application of the method; ***selecting a goal*** again, if necessary.	***recognition*** of the existence of a problem; ***diagnosis***, clarifying and defining the issues; ***(a) search*** for ready-made solution ***or (b) design*** of a custom-made solution; ***evaluation or choice***, selecting the preferred solution; ***authorization***, a higher organizational level sanctioning implemen-tation.	***definition*** of the problem; ***description*** of problem in 4 dimensions: identity, location, timing, and magnitude; extraction of key information in these 4 dimensions to ***generate possible causes***; ***testing*** for most probable cause; ***verification*** of the true cause; ***resolution*** (can coincide with verification if through verification problem is corrected).	***recognition*** that decision has to be made; ***formulation*** of decision problem in terms of objectives; ***alternative generation*** or identification of all possible courses of action; ***information search*** regarding each course of action and likely outcomes; ***evaluation*** of ***and choice*** between alternatives; ***action*** upon decision made, receiving ***feedback*** about the outcomes.	***identification*** or realization of fact that something is out of order; ***definition*** of the problem as discrepancy between desired and actual state; ***diagnosis*** of problem's cause(s); ***generation of alternatives*** i.e. measures to counteract causes; ***evaluation of alternatives***, weighing pros and cons of the various solutions; ***choice / action***, choosing and implementing preferred alternative.

Table 1 continued

Dörner & Wearing (1995) action regulation model	**Noble (1989) situation assessment (SA) model**	**Beach & Mitchell (1987) image theory**	**Montgomery (1983) dominance search model**	**Klein (1993, 1997) recognition primed decision (RPD) model**	**Cohen & Freeman (1997) recognition / metacognition (R/M) model & STEP tool**
goal-elaboration; information collection and formation of hypothesis about the reality; ***prognosing*** what will happen; ***planning*** a course of actions ***and decision making***; ***monitoring*** effects of the implemented actions; ***self-reflection***.	***test process*** where expectations are compared with observations, through a combination of internal input (knowledge) and external input (data about situation and context); ***create or update process***, where (a) the situation representation is refined is or (b) a new one created.	For **adoption decisions**, concerning addition of goals and plans to the agenda: ***compatibility test*** to screen out unacceptable goals and plans; if more candidate plans remain, ***profitability test*** to select the best. For **progress decisions** used to decide if a plan is reaching its objectives: ***compatibility test***; ***decision*** either to stay with status quo or to change plan or goal.	***pre-editing***, or selecting and using criteria to screen alternatives; ***finding*** a promising ***alternative***; ***testing if*** this is the ***dominant*** alternative, at least as attractive as competing alternatives on relevant criteria and exceeding every alternative on at least one criterium. (a) if so, ***decision*** for this alter-native follows (b) if not, ***dominance (re)structuring*** follows.	***experiencing the situation***; only if unclear, ***diagnosis*** follows; ***recognition*** of situation (goals obvious, cues attended to, expectations formed, and course of action recognized); in more complex cases, ***evaluation*** of action via mental simulation; after which, if necessary, ***modification*** of the action plan follows, or a return to evaluation of other possible actions; ***implementation of action***.	create a ***Story*** to explain the situation; ***Test*** for conflicts in the story and patch it up where needed; ***Evaluate*** the resulting story and if necessary go back to first step; develop contingency ***Plans*** for the weakest assumptions in the story.

The most obvious common characteristic of these models is that they divide problem solving or decision making processes into several phases. All models include a phase early on in the process aimed at understanding the problem. Most of the models, except Noble's situation assessment model, also include a phase in which an action plan is devised or chosen from an existing repertoire. The phase in which these actions are actually implemented is included in even fewer models. Next to this discrepancy between the models, there are several other differences. For example, what is treated as one phase in one model can be divided over several phases in another. Some phases are not even included in all of the models, not only the implementation phase as mentioned before, but also the phase in which people recognize or detect that there is a problem, the phase where the problem's causes are identified, and the phase where the implemented actions are evaluated.

These differences can in part be explained by the researchers' definitions of what constitutes problem solving. Frensch & Funke (1995) for example concluded that these definitions differ not only with regard to which actions and how many events are classified as problem solving, but also with regard to how fuzzy the boundaries are between what is considered problem solving and what not. Some researchers prefer to label their research as decision-making research instead of research on problem solving. In the strictest sense, where decision making is concerned, the actual implementation of what has been decided and evaluation thereof would be outside the scope of the process. However, many of those involved in decision making research have considered these steps as part of the process, thus causing the two research domains to overlap each other to a large extent.

Another part of the explanation for the differences between the models lies in the situation they were developed for. One of the features used by problem solving and decision making researchers to describe their typical area of interest, are the characteristics of the problem: how structured is the problem, how isolated from or interwoven with other problems or processes (e.g. Orasanu & Connolly, 1993; Meacham & Cooney Emont, 1989), in which task did the problem occur (e.g. Frensch & Funke, 1995), and which kind and number of solution(s) are applicable (Meacham & Cooney Emont, 1989). Furthermore, with regard to these characteristics of the problem, Getzels (1982) distinguished three types of problems: *presented* problems which are created, discovered and presented by others to the problem solvers, *discovered* problems which are created by others and discovered and solved by the problem solvers, and *created* problems which are created, discovered and solved by the problem solvers. A detection phase, for example, is not relevant for presented problems. In addition, if the problem structure is clear already at the moment a person detects the problem's existence, the need for efforts aimed at understanding the problem decreases.

With regard to planning and implementing solutions, the availability of standard solutions may decrease the need for efforts aimed at these goals.

Other features used to describe or characterize problem solving research interests are the characteristics of the problem solver(s), and the setting, or circumstances under which the problem has to be solved (Frensch & Funke, 1995; Orasanu & Connolly, 1993): characteristics of the environment (e.g. stability), the available resources (people, knowledge, strategies, time, tools and equipment), and stakeholders.

Given this wide variety in problems, problem solvers, and problem contexts, it is only logical that the applicability of any given phase model is limited to a certain area. The models from Klein and from Cohen and Freeman for example stem from a research stream that originated in the late 1980s, known as Naturalistic Decision Making (NDM) (e.g. Klein, Orasanu, Calderwood, & Zsambok, 1993; Zsambok & Klein, 1997; Flin, Salas, Strub, & Martin, 1997; Salas & Klein, 2001; Pliske & Klein, 2003). Naturalistic decision settings are characterized by uncertain, changing environments, in which organizational norms have to be taken into account, multiple players are involved in decision making, the stakes and also time stress are often high, problems are not well-structured, goals may be unclear, shifting or even competing, and multiple actions are required over time, producing feedback in between. The NDM researchers found that many of the insights in decision making processes stemming from laboratory experiments do not translate very well to the settings described above, and chose to study decision making in real life, naturalistic settings instead, with all the difficulties associated with doing research in uncontrolled environments. Many of their reservations and principles also apply to recovery processes taking place in work environments.

An additional difference between the phase models presented in figure 1 is whether they were meant to be descriptive or prescriptive, and how flexible they are in accommodating different processes, where some phases may not occur at all and some may be repeated at a later stage. Cohen and Freeman's STEP model is prescriptive, the idea is that performance in decision tasks involving perceptual input can be improved by following the STEP procedure. The prescriptive validity of their model has been confirmed in several studies (Cohen and Freeman, 1997). The phase models proposed by Dörner and Wearing, by Newell and Simon, by Mintzberg et al., by Noble, by Beach and Mitchell, by Montgomery, and by Klein are examples of descriptive models, they describe the problem solving process without insisting that that is the best or only way. In fact, Mintzberg et al. even disconfirmed the descriptive validity of their model, they found that none of the strategic decision processes they studied smoothly followed the order specified by their model. On the other hand, by differentiating between steps (at which moment, where in the sequence of the process was the action performed) and phases (what was done, what was the goal) of a problem solving process, Lipshitz and Bar-Ilan (1996) confirmed the descriptive validity of their model (i.e. with regard to the order in which phases occur). They disconfirmed the prescriptive validity of their model, finding no significant differences in phase sequence between successful and unsuccessful problem solving processes. They also found a tendency to skip forward in the model, directly to action. This is an indication that a process model, to accurately describe possible problem solving processes, should be flexible with regard to transitions between phases. Dörner and Wearing (1995) also noted that different sequences through their phase model and iterations are possible, for example when collected information sheds a new light on goals. The models proposed by Newell and Simon, by Mintzberg et al., by Noble, by Beach and Mitchell, by Montgomery, by Klein, and by Cohen and Freeman all are flexible to some extent:

- either with regard to the specific form a phase takes (Mintzberg at al. model, Noble's model; as can be seen from the options (a) or (b) given in certain phases),
- or the actual occurrence of phases (Klein's model, Montgomery's model, Beach and Mitchell's model; as can be noted from words like 'if...', 'in ... cases'),
- or the amount of times a sequence of phases is repeated (Newell and Simon's model),
- or the sequence in which phases take place (Cohen and Freeman's model allows for looping back and so do Klein's model and Montgomery's model; see words like 'go back', 'return', 'start over').

The idea of recovery processes consisting of a sequence of process phases was also found in the literature available from the domain of human reliability and safety on recovery from failures. As can be seen from table 2, general agreement exists among researchers in this domain with regard to the existence of at least three recovery process phases.

A few of the researchers listed in table 2 have included a phase before the actual detection that something has gone wrong: the occurrence of an error (Zapf & Reason, 1994); the emergence of a mismatch (Rizzo, Ferrante, & Bagnara, 1995); or the presence of threats, which can lead to errors (Bove, 2002; Bove & Andersen, 2001). From the detection phase onwards, the same phases are distinguished by most researchers even though sometimes different names are used: a phase during which people try to find out what exactly has gone wrong and how this could happen, and a phase where they plan and implement countermeasures.

As for the flexibility of the models presented in table 2, some of them cover linear processes only, while others allow for alternative sequences, iterations, and skipping phases. Many of the researchers listed in table 2 (e.g. Frese, Brodbeck, Zapf, & Prümper, 1990; Frese, 1991, Zapf & Reason, 1994; Sellen, 1994; Kontogiannis, 1999) have explicitly stated that by necessity, the detection phase always has to occur before the rest of a recovery process can take place. After all, no further recovery steps can be initiated for undetected deviations. Johannsen's fault management process model (1988) is not flexible with regard to both the order and the actual occurrence of the three phases. Embrey and Lucas (1988) did not specify anything at all with regard to the order in which the phases take place and the need for all phases to occur, and, except for stating that detection has to occur before the rest of the recovery process, neither did Zapf and Reason (1994), nor Sellen (1994). Bagnara, Ferrante, Rizzo, and Stablum (1988) noticed differences in their studies with regard to the analysis phase and incorporated these in their model: the amount of cognitive processing can vary, and sometimes for various reasons no analysis is performed at all. Similarly, Frese et al. (1990) and later again Frese (1991) pointed out that not every process step has to be done, especially localization and explanation of the error may not be necessary. They noted that different parts of the error handling process might be intermingled, too. Rizzo et al. (1995) also found that recovery is not necessarily based on a full understanding of the mismatch and error. In addition, they noted that in some cases, the mismatch emergence, detection and identification phase overlap. Van der Schaaf (1988, 1995) also provided evidence that the more analytic localization phase does not necessarily always take place. Kontogiannis (1999) proposed that the need for an explanation phase depends on the time available and the need to know the causes before corrective actions can be planned. Bove (2002) and Bove & Andersen (2001) even decided not to incorporate an explanation phase in their model.

Table 2 Recovery process phase models from the domain of human reliability and safety

	Johannsen (1988)	Embrey & Lucas (1988)	Bagnara, Ferrante, Rizzo, & Stablum (1988)	Frese, Brodbeck, Zapf, & Prümper (1990)	Frese (1991)
phase 0					
phase 1	detection (discovering undesired deviations)	detection	detection	detection (knowing that an error has occurred)	detection (knowing that an error has occurred)
phase 2	diagnosis (identification of causes of deviation, i.e. locations and types of failure)	diagnosis (of abnormal state)	causal analysis (automatic, conscious, or exploratory; sometimes not necessary or skipped)	localization (being able to explain the error)	explanation (knowing how and why the error came about)
phase 3	correction (return to normal state) or compensation (return to at least an acceptable state)	correction	correction (or overcoming the mismatch in another way)	recovery (corrections)	handling (imple-menting the corrections)

Table 2 continued

	Zapf & Reason (1994)	Sellen (1994)	Rizzo, Ferrante, & Bagnara (1995)	Van der Schaaf (1995)	Kontogiannis (1997, 1999)	Bove (2002), Bove & Andersen (2001)
phase 0	error occurrence		emergence of mismatch (between feedback and expectations or assumptions)			threats, which can lead to errors
phase 1	detection	detection (knowing that an error has occurred)	detection (awareness of what is wrong)	detection (of deviation, symptoms)	detection (of a mismatch between observations and expectations)	detection
phase 2	explanation (together with detection part of diagnosis phase)	identification (knowing what was done wrong)	identification (causal assessment regarding the mismatch)	localization (of the causes of this deviation)	explanation (localization, where and why did the error occur)	-
phase 3	recovery (planning and execution of corrective actions; often plan may already be at hand)	recovery (knowing how to undo the effects of the error and return to desired state)	recovery (overcoming the mismatch)	correction (to return to the normal state)	correction (re-assess situation, develop plan and execute)	correction consisting of response *(trap, exacerbate, or fail to respond)* and outcome *(recovery i.e. inconsequential, undesired state, or additional error)*

What lessons can be learnt from this review of existing phase models describing processes followed by people to make decisions, solve problems, and recover from errors or other failures? First of all, recovery processes seem to consist of several process phases, and different goals and thus different types of actions and/or cognitive processes are associated with each phase. Second, while detection is always the first phase to occur, what actually is detected at the start of a recovery process can either be the failure itself, or initially only the resulting deviation or problem (the symptoms), of which the causes may not immediately be obvious. Third, after the detection phase, not all recovery processes involve the same steps or phases. A recovery process model has to be flexible with regard to the phases following detection. In some cases, for example, it may not be necessary to identify what caused the deviation before successful corrective actions can be taken. Iterations of the phases after detection may also occur, leading to recurrence of phases in later steps in the process. In some cases, the people involved may decide, either immediately after detection or after some efforts aimed at further explaining the detected deviation, not to implement any countermeasures. These lessons have shown that a generally applicable recovery process phase model can be descriptive at best, no fixed, optimal, universally applicable recovery process can be prescribed; these processes are too situation-specific. Finally, even within a process phase sometimes different sub-phases can be distinguished, or the exact goal people strive for within a certain phase may differ depending on the circumstances. Some quick, temporary countermeasures may need to be implemented, for example, before further, more permanent recovery actions can be planned and taken.

Based on these insights, I propose the following phase model to describe the different types of possible steps (i.e. actions and/or cognitive processes), and transitions involved in recovery processes.
First (and in this thesis treated as part of the recovery process, not merely the trigger):

- Detection of the (combination of) failures or at least the immediately resulting deviation or problem situation (D).

Followed by any combination of (including entirely skipping one or both of these phases, but also recurrences):

- Explanation of the deviation, how it has manifested itself, and what its causes were (E). In the literature (cf. table 2), this phase is also referred to as localization, diagnosis, analysis, or identification.
- Countermeasures aimed at returning to the normal situation or at least limiting the consequences (C). Also for this phase, many names were found in the literature (cf. table 2), such as correction, compensation, handling, or even recovery, which in this thesis refers to the entire process and not merely one phase.

Figure 4 contains a graphical representation of the proposed phase model.

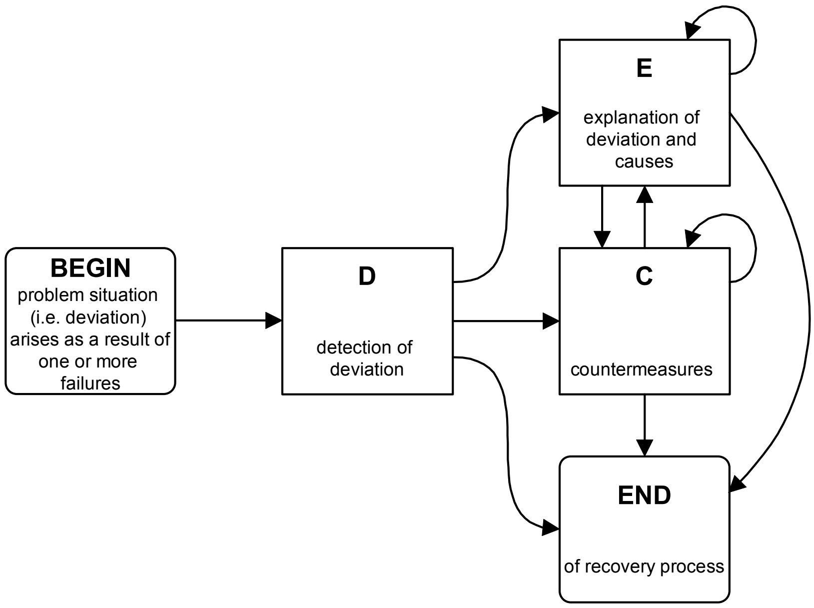


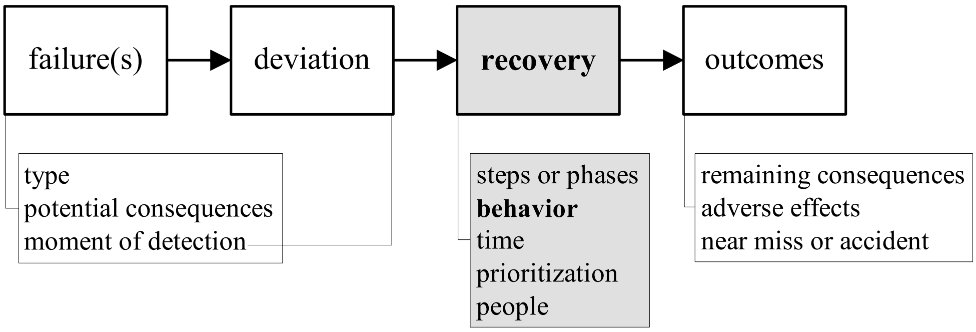
Figure 4 Proposed recovery process phase model

Most of the existing research into recovery processes has focused on how different types of failures or the immediately resulting deviations are detected (e.g. Sellen, 1994; Zapf, Maier, Rappenberger & Irmer, 1994; Doireau, Wioland & Amalberti, 1997; Sarter & Alexander, 2000; Rizzo, Bagnara & Visciola, 1987; Allwood, 1984; Rabbitt, 1978; Mason & Redmon, 1992; Curry & Ephrath, 1976; Curry & Gai, 1976; Wickens & Kessel, 1981; Curry, 1981; Moray, 1981). The insights from the domains of signal detection, interface design and alarm management also add to our understanding of the detection phase. The attention given to detection is well deserved, since no corrective actions will be initiated for failures that remain undetected. On the other hand, much less is known about the actions and cognitive processes that follow the detection phase. The main contributions to our understanding of these later parts of the recovery process stem from the research on fault finding and fault diagnosis (e.g. Duncan, 1981; Marshall & Shepherd, 1981; Patrick & Stammers; 1981; Duncan, 1987a, 1987b; Morrison & Duncan, 1988; Brinkman, 1990; Kostopoulou & Duncan, 2001). This research, however, mainly focused on locating the (technical) origins of faults in technical systems, a limitation that is addressed in more detail in section 2.3. In order to add to these already existing insights, in the research described in this thesis no restrictions were made with regard to which recovery process phases were studied and with regard to the types and combinations of failures (see section 2.3) preceding the recovery process.

To verify the proposed recovery process phase model, to establish which recovery steps or process phases and transitions between those actually occur during recovery processes in practical settings, and how based on this different types of processes can be distinguished, and to develop a more detailed insight in the goals of each of the phases, I formulated

Research question 1:
a) With regard to the steps and transitions involved, does the proposed recovery process phase model accurately describe the processes people follow to recover from failures in work situations,
b) which different types of recovery processes can be distinguished based on which phases occur when during the process, and
c) which additional details, in the form of subgoals, can be distinguished for each of the phases of the proposed phase model?

2.2.2 Behavior involved in recovery steps



Recovery process steps do not only serve one of the three goals distinguished in the previous subsection, that is, detection, explanation, or countermeasures, they can also be characterized by the behavior and amount of cognitive processing they require. Looking from an organization's perspective at people's behavior during recovery processes, a distinction can be made between planned and unplanned recovery steps.

Planned recovery steps involve the activation of defenses or barriers that are built into the organization to avoid negative consequences (Svenson, 1991, 2001; Hollnagel, 1999). Examples of barriers or defenses are automatic safety controls, or procedures to follow under certain conditions. Such procedures can be formally documented rules or work instructions, as a result of an organization's need to formalize behavior (e.g. Mintzberg, 1983), or they can be undocumented. In this thesis, also unwritten work practices (cf. what for example Brief & Downey, 1983, described as implicit rules) are considered to be built into the organization and thus qualify as planned behavior, as long as they are generally accepted and followed by all people involved in the associated tasks. If a person's recovery behavior is based on training this person has had in fault diagnosis (e.g. Patrick & Stammers, 1981; Duncan, 1981, 1987a; Kostopoulou & Duncan, 2001), this also qualifies as planned behavior.

Unplanned steps, on the other hand, most often have a more ad-hoc nature and depend on the creative thinking and problem solving abilities of the people involved. Even though a person who performs such unplanned recovery steps may have performed these steps before, as long as this behavior is not a standard practice within the organization, known to everyone in the same job, these steps still qualify as unplanned behavior. The word 'unplanned' for these steps does not mean that the steps are performed without any plan to guide them, it merely indicates that the plan for these steps is neither provided through explicitly documented procedures nor through implicit procedures. If a plan is needed for such steps, it is devised by the person(s) involved.

Some recovery steps may actually involve a mixture of planned and unplanned behavior; for example in unforeseen situations where people are to some extent still assisted in recovery steps by the automated systems or other equipment they work with, while having to use these planned provisions in novel, unplanned ways.

In existing research, the distinction between planned and unplanned recovery steps has been made, for example, by Mintzberg et al. (1976), who differentiated between ready-made and custom-made solutions in problem solving, that is, between planned and unplanned countermeasures. Furthermore, with regard to the detection phase, Allwood and Montgomery (1982), for example, in a study of how people detect errors in a statistical problem solving task, differentiated between unplanned or spontaneous detection, and planned detection triggered by explicit instructions.

While most of the existing research on recovery has focused on barriers and defenses, i.e. (pre-)planned recovery provisions, in this thesis special attention is also given to ad-hoc steps. This is where people's creative abilities are most visible. The terms unplanned and ad-hoc refer to the same types of spontaneous actions or cognitive processes and are used interchangeably in this thesis.

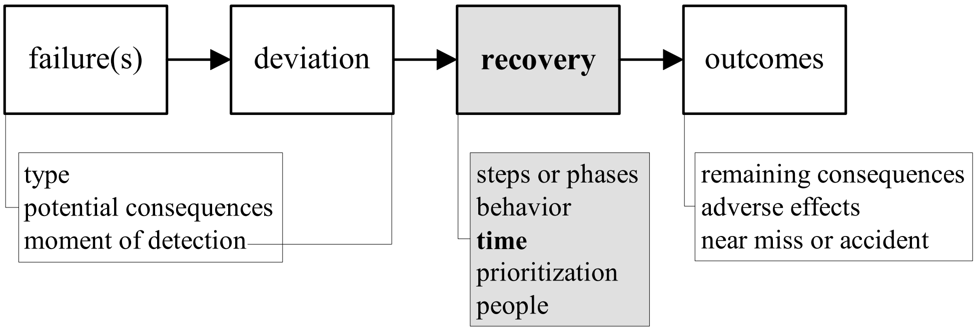
Related to the difference between pre-planned and ad-hoc steps is the question whether the recovery steps involve skill-based, rule-based, or knowledge-based behavior. This is a categorization proposed by Rasmussen (1976, 1986) and is widely used in the domain of human reliability (e.g. Reason, 1990) to categorize levels of behavior or performance, during which different types of errors can occur. Johannsen (1988) demonstrated that the categorization can also be used to describe different levels of operator behavior during fault management tasks. The categorization is based on the amount of conscious reasoning or information processing involved. Skill-based behavior is sort of automatic, and requires hardly any thinking. It involves sensorimotor performance, which is triggered by an intention and for which no further conscious control is needed. Examples of skill-based performance are walking, or riding a bicycle, that is, as long as the person performing these activities has progressed past the learning stage for these activities, has had plenty of practice, and does not encounter any complicating factors during performance. Rule-based behavior requires a little more reasoning, since the appropriate action that corresponds with the encountered condition or situation needs to be selected first from a known repertoire of predefined actions. These actions are still very familiar for the person performing them, and are structured and controlled via stored rules derived from previous experience or instructions. An example of rule-based performance is that of a person following a familiar recipe to bake a cake, whereby the recipe tells this person which quantities of flour, butter, eggs and sugar to use, in which order to mix them, what temperature the oven should be, how long the cake should stay in the oven, and so on. Separate actions in this procedure may actually be performed on a skill-based level, such as mixing the ingredients, but the overall activity is governed by rule-based behavior. Finally, knowledge-based behavior is applicable for novel, unfamiliar situations and involves a substantial amount of reasoning. During knowledge-based behavior people formulate their goals based on an analysis of the situation and their overall aims. To this end, based on stored knowledge and the results of the situation analysis, a person develops a ‘mental model’ of the situation as a guide for further actions. People who are learning a new task, either at work (e.g. learning to replace filter in a pump), school (e.g. learning to multiply fractions), or in everyday life (e.g. learning to drive a car), engage in knowledge-based performance. Over time, after enough practice, the behavior level required for them to perform these tasks may change to rule-based or even skill-based. Basically, anyone entering an unfamiliar situation that requires him or her to act correspondingly will engage in knowledge-based behavior.

The level of information processing behavior involved when a person is performing a task also has consequences for how this performance can be supported optimally. Only during knowledge-based behavior, people are actively looking for feedback to control their performance, but also at the other behavior levels it may be vital that important cues get through to the people involved. I propose that all this equally holds for recovery steps and want to explore which types of recovery steps involve which behavior level.

To find out to what extent different recovery steps are pre-planned by the organization and which role the different information processing behavior levels play, I posed

Research question 2:
For each of the distinguished types of recovery process phases (i.e. detection, explanation, and countermeasures),
a) which proportion of all the steps belonging to that phase involves ad-hoc behavior and which proportion pre-planned behavior, and,
b) furthermore, which proportion involves skill-based behavior, which proportion rule-based behavior, and which proportion knowledge-based behavior?

2.2.3 Time spent on recovery steps and process



In a society where most actions and cognitive processes have to be completed under some form of time pressure, one may expect that this also applies to recovery processes, especially when negative consequences can occur soon and other tasks compete for attention. Deviations are hardly ever planned, and the corresponding recovery steps have to be fitted in with ongoing tasks. In his explanation how actions and corresponding mental processes are embedded in time, Hollnagel (2002) also referred to the presence of time constraints. He pointed out that limited time is available to evaluate events, decide what to do, and to actually implement the planned actions, and that in complex processes goals may even change over time, for example on the basis of new information that has become available.

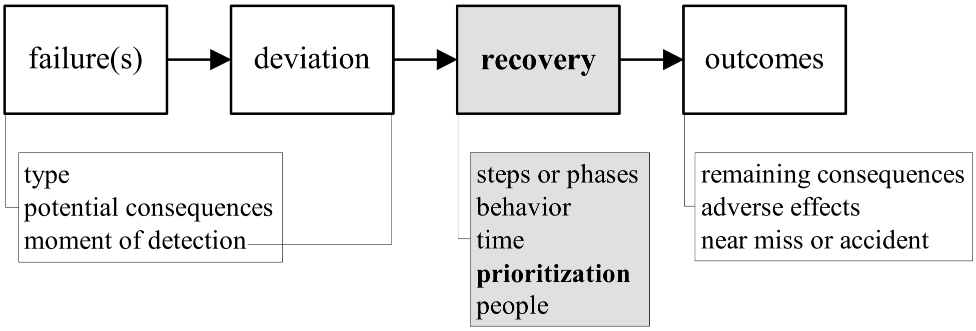
Time pressure can be the result of what Rasmussen (1997) referred to as the migration of activities to the boundaries that govern performance in an organization operating in our current, dynamic society. Employees do not only have to operate within the boundary to economic failure, but may actually be pressured by management towards working more efficiently, as staying further away from this economic boundary means an advantage over competitors. This drive towards efficiency can lead to an increase in time pressure perceived by the employees. Employees also have to operate within their own limitations, within the boundary to unacceptable workload, and may start to look for ways to minimize the amount of effort and time they need to spend to get a job done. Finally, employees have to operate within the boundary of acceptable performance as well, since crossing this boundary involves safety and reliability risks. Efforts spent on safety and reliability may have moved this boundary towards being more restrictive, but as a result of management pressure towards efficiency and the desire to minimize one's workload, employees may be tempted to cut corners, and occasionally cross this more restrictive boundary of acceptable performance, and come near the more distant, absolute boundary. In doing so, they may go too far and create a deviation from which they need to recover (if still possible), thus adding additional tasks to their workload, which can also increase the time pressure they are under.

In the domain of human-machine interaction, the term 'supervisory control' is used to describe a person's interaction with 'intelligent' automated systems. Important topics in research regarding supervisory control (e.g. Sheridan, 1997) are where in the control loop of a human-machine system the human operator should best be placed, how much control this operator should have, and what feedback to give when to the operator, and how. The amount of time a human operator has available to spend on recovery steps, if these are needed, and possibly on waiting for feedback in between those steps, may depend on his role and position in the control loop, and also on the timing, nature and amount of feedback this operator receives.

To find out how much time people spend on separate recovery steps, given their position in a control loop, and in spite of or as a result of the time pressure and other pressures they experience, and how much time passes by before all the recovery steps are completed, I formulated

> *Research question 3:*
> *How long do*
> *a) different types of recovery steps at various stages in the process, and*
> *b) the entire process*
> *take?*

2.2.4 Prioritization

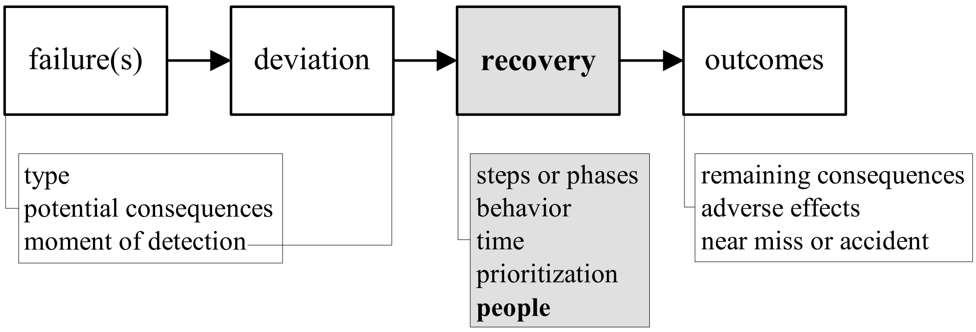


Closely related to the amount of time spent on recovery steps is how priorities for these steps are established. Task prioritization is one of the adaptive responses to non-routine events distinguished by Waller (1999) in her study of airline crew behavior. According to goal theory (e.g. Geen, 1995), actions towards stronger, more urgent goals get precedence over actions towards other goals. Because of the potential outcomes if no recovery would take place, I expect the goal to recover from a dangerous situation to be more compelling in most cases than any other goals a person may be entertaining at that moment. For people to give such priority to recovery from a deviation, they have to perceive this deviation as being dangerous and the potential consequences as something they want to avoid. In section 2.3, this issue is discussed in further detail.

In the literature on time management, not only urgency but also importance are put forward as factors that affect prioritization (e.g. Covey, Merrill & Merrill, 1994; Koch & Kleinmann, 2002; Claessens, Van Eerde, Rutte, & Roe, 2003). Furthermore, action theory research (e.g. Heckhausen & Kuhl, 1985) not only mentions urgency and importance as factors in ‘wants’ being transformed into intentions and intentions into actions, but also time, means, and opportunity. The factor opportunity seems less relevant at a stage where a deviation has already led to an opportunity for recovery. As for the factor means, in some cases certain specific materials, tools or equipment may be necessary before the recovery steps can be performed, and in some cases also the assistance of others. Lacking the required means may lead to postponing recovery steps. Regarding the factor time, even though the necessary recovery tasks may be urgent and important, there may be cases where other tasks also require immediate attention, which may influence the available time and the priorities. Another role the factor time may have in prioritization lies in the amount of time needed for recovery steps. There is a problem if this is more than the available time (cf. Hollnagel, 2002), both with respect to other tasks needing attention, and with respect to the first possible occurrence of negative consequences. In the case of recovery process steps, the time available before the potential negative consequences can occur corresponds with the factor urgency. The type, severity, and likelihood of the potential consequences (see section 2.3) correspond with the factor importance. In summary, both the presence of other tasks requiring immediate attention, the time needed for recovery, the availability of the necessary means, and the urgency and importance of recovery steps may be important factors in prioritizing tasks when people are confronted with a situation in need of recovery. To investigate how people prioritize their tasks when recovery from a deviation is needed, I posed

> *Research question 4:*
> *Based on which factors do people prioritize tasks when a deviation needing recovery presents itself at work?*

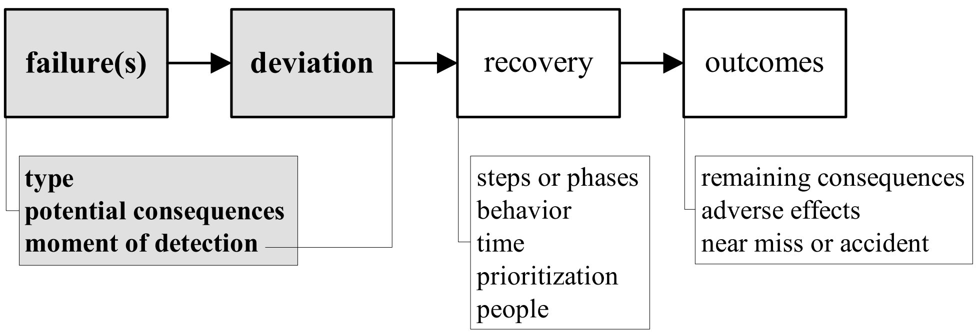
2.2.5 People involved in recovery



Related to what was discussed in subsections 2.2.3 and 2.2.4 with regard to time and prioritization is the question which and how many people are involved in a recovery process. Recovery processes are not necessarily one-person efforts. Doireau et al. (1997) demonstrated the important role of people other than the person who committed the error in the detection of errors. Both Hutchins (1996) and Wioland & Amalberti (1998) argued that detection by others depends on the amount of context sharing between the people committing - and the people detecting the error. Involvement of others may also be important at later stages in the process. After all, task re-distribution, sharing, or handing over tasks to others are some of the possible ways to deal with a sudden increase in workload involving tasks for which time pressure is high, that is, tasks with high priority. Waller (1997), for example, referred to the possibility of (re-)allocating (sub)tasks to other group members in multitask situations. This possibility includes both collaboration with and delegation to others. Maule (1997) also mentioned delegation of work to others as one of the modes to adapt to time pressure. Similarly, Kecklund and Svenson (1997) found that operators used delegation of tasks to others as a way to cope with demanding work situations. Factors similar to those that are expected to influence task prioritization are also likely to play a role in the decision to involve others. If recovery steps are important and urgent enough that they cannot wait, there are several types of resource- (i.e. means and time) related reasons possible to involve others: people already involved may not have the required expertise and skills (cf. Waller, 1997), or (cf. subsection 2.2.4 and Heckhausen & Kuhl, 1985) they may not have the necessary tools and materials, or enough time, or enough hands or manpower without involving others; but on the other hand, others who can assist (i.e. additional resources) may not always be available. The reason for involving others may also be rule- or procedure-related, that is, people may not be allowed to perform tasks that are officially the responsibility of others. Also implicit 'rules' (cf. subsection 2.2.2) may influence the decision to involve others, in some organizations it may not be considered acceptable to ask for help, or to warn others, or comment on their actions, whereas in other organizations such behavior may actually be encouraged. Multiple reasons for involving others may apply at the same time. To extend existing insights in the role different people can play in recovery to the entire recovery process I formulated

Research question 5:
If others than those already involved become involved in a recovery process at work:
a) what are the organizational roles of these people,
b) when do they become involved,
c) for what reasons do they become involved, and
d) what is the effect of their involvement?
Furthermore, if those involved in detection do not involve any others in the rest of the recovery process at all, for what reasons does this happen?

2.3 Events leading up to recovery processes



Generally speaking, the need for recovery arises when a failure or combination of failures has occurred that has produced a deviation with the potential to lead to negative consequences. These failures and resulting deviations do not only precede and necessitate the recovery process that follows, they may actually even influence the form this recovery process takes as well. This section starts by taking a closer look at the concept of deviations, followed by a discussion of the different underlying failure types that can be distinguished, and their role in incident causation. The attention then shifts to the potential negative consequences of the deviations, and after that to the moment the deviations are detected. The available insights and gaps in knowledge regarding the influences on a recovery process of all of the issues addressed are also discussed throughout this section.

What should be noted is that not all variations from normal, or operational anomalies (referred to as 'surprises' by Koornneef, 2000) are considered deviations requiring recovery. Which variations are acceptable and which not may depend on various factors, such as the type of process one is controlling (e.g. technical production system or human behavior), control limits set by the organization, procedures, social norms and culture, legislation, and of course potential consequences as well (see later in this section). The threshold between acceptable and unacceptable variations may even be dynamic (see also Hollnagel, 2002) – depending on the circumstances under which a variation takes place. During high workload situations people may set this threshold a little higher than during low workload. Furthermore, a difference can exist between the actual situation and a person's perception of that situation. That is, people may incorrectly perceive a situation as a deviation and initiate recovery steps that are not really needed. Or on the other hand, actual deviations may not be perceived as such, thus missing an opportunity for recovery.

Deviations (i.e. variations from the normal, intended situation that are considered unacceptable) can most often be traced back to a number of underlying causal factors. The purpose of incident investigation is to identify such causal factors, and to establish how they led to a deviation, and, if applicable, from there to negative consequences.

As Reason (1991) argued, over the years the scope of incident investigation has changed substantially, and so have the models of incident causation to which incident investigators adhere. By means of examples from the domains of rail transport and nuclear power generation, Reason (1991) illustrated how in the early stages of any new technological development incident investigations and the preferred preventive measures focus mainly on technical failure as the cause of incidents. He went on to explain how at a later stage in such developments the focus usually switches to human error (i.e. human failures) and human-machine mismatches, and finally to organizational factors. Since this distinction between technical, human, and organizational failures appeared to work well before in several studies into the causes of near misses and more serious incidents (e.g. Van der Schaaf, 1992, 1995, 1996; Van Vuuren & Van der Schaaf, 1995; Van Vuuren, 1998, 1999), the decision was made to also use this distinction in the research described in this thesis. The point of reference that was used to differentiate between the three types of failure was that of the people at the operational level in an organization. Otherwise, for example, also technical failures caused by errors made by the designer and organizational failures created by bad management decisions would qualify as human error, even though these human errors are outside the control of those confronted with the resulting deviations. The following definitions were used for the different failure types:

Technical failures are failures related to the design or the construction of technical systems, installations, machinery, tools and equipment, or simple material or component defects (Van der Schaaf, 1992).

Human failures or errors are those failures committed by the people at operational levels in the organization. Several different types or subcategories of errors can be distinguished. As Reason (1990, page 10) stated with regard to the categorization or classification of errors: “A taxonomy is usually made for a specific purpose, and no single scheme is likely to satisfy all needs. Nearly everyone who has published in this field has devised some form of error classification.” Working towards a model of errors himself, too, Reason (1990) distinguished three cognitive stages involved in an action sequence, namely planning, storage of the plan, and execution of the plan. Specific error types are associated with each of these stages, respectively mistakes, lapses, and slips. These error types can be linked with a specific level of behavior or performance at which they can occur: skill-based, rule-based, or knowledge-base performance (Rasmussen, 1976, 1986; Reason, 1990; see also subsection 2.2.2). Slips and lapses are skill-based errors, mistakes can occur either during rule-based or knowledge-based performance. Many other researchers in the domain of safety and reliability have also used this distinction between skill-based, rule-based, and knowledge-based errors, for example Hale and Glendon (1987) in their model of behavior in the face of danger, and Van der Schaaf (1992, 1996), who used it as a departing point for the development of a more detailed taxonomy of failure types. Most handbooks in the field of safety, reliability, human factors, and ergonomics include a discussion of Rasmussen’s (1976) model of skill-based, rule-based, and knowledge-based performance (the SRK model), e.g. Wickens, Gordon, & Liu (1998). Even outside academic circles, as for example Van der Schaaf (1992) noted, the use of this model is widespread and its concepts are easily understood and accepted. Therefore, also in the research for this thesis, within the category of human failures, a distinction was made between skill-based, rule-based, or knowledge-based errors.

Irrespective of whether they were based on the SRK model or not, many error taxonomies have provided an even more detailed subdivision of error types. A well-known example is Norman's (1981) categorization of action slips (i.e. unintended actions), based on the notion that action sequences are controlled by sensorimotor knowledge structures referred to as 'schemas', and that several types of slips can occur at different stages in such action sequences. The main categories of slips Norman (1981) distinguished are errors in intention formation, faulty activation of action schemas, and faulty triggering of active schemas.

Finally, organizational failures are made at management levels in the organization. Such failures most often indirectly lead to incidents by triggering other failures (Van Vuuren, 1998, 1999; Reason, 1997). They can be related to the organizational (safety) culture, to the organization's strategy and goals as expressed in management priorities, or to the organization's structure reflected in the quality and availability of protocols, procedures or instructions, or in how the transfer of required knowledge and information to employees is established (Van Vuuren, 1998, 1999).

An in-depth way to analyze an incident, and thus to establish its causes, is to perform a fault tree analysis (e.g. Kirwan, 1994; Henley & Kumamoto, 1981) or a causal tree analysis (Van Vuuren & Van der Schaaf, 1995). Causal tree analysis is based on the fault tree approach, but also allows for the inclusion of information about the recovery steps taken (if any). Fault- and causal tree analyses aim at systematically tracing how the incident was caused by a number of factors which each on their turn have their own causes, and so forth, until the so-called root causes of the incident are identified, for which no further explanation can be found within the organization or context in which the incident took place. These root causes form the bottom of the causal tree and can be categorized according to their nature (i.e. technical, human, or organizational failures). The use of these techniques and similar ones have demonstrated that most incidents have multiple root causes, it is hardly ever just one failure factor on its own that leads to an incident (e.g. Reason, 1990; Wagenaar & Groeneweg, 1987).

While many different types of failures or combinations thereof can lead to the need for recovery, most of the existing research on recovery processes has focused on recovery from errors, that is human failures (e.g. Rabbitt, 1978; Norman, 1981; Allwood, 1984; Embrey & Lucas, 1988; Bagnara et al., 1988; Frese et al., 1990; Mason & Redmon, 1992; Brodbeck, Zapf, Prümper & Frese, 1993; Sellen, 1994; Zapf et al., 1994; Rizzo et al., 1995; Edmondson, 1996; Doireau et al., 1997; Kontogiannis, 1997 & 1999; Sarter & Alexander, 2000; Bove, 2002). Furthermore, as was pointed out already in subsection 2.2.1, from all the recovery process phases, most of the attention in this research has been given to the detection phase. Existing research on fault finding and fault diagnosis (e.g. Duncan, 1981;

Marshall & Shepherd, 1981; Patrick & Stammers; 1981; Duncan, 1987a, 1987b; Morrison & Duncan, 1988; Brinkman, 1990; Kostopoulou & Duncan, 2001), as was also mentioned in 2.2.1, mainly focused on locating the (technical) origins of faults in technical systems. This research was more concerned with protocols, training and support for the steps involved in fault diagnosis than with the effect these faults had on the detection, diagnosis and correction that followed. The only models in table 2 (see subsection 2.2.1) developed to describe recovery from several types of failures are those from Van der Schaaf (1995) and Johannsen (1988). These authors, however, did not provide any research data either on how these failure types actually influence the recovery process.

As a result, the only insights available so far on how failures impact the recovery process that follows focus on the differences in recovery between skill-based slips and lapses, and rule- and knowledge-based mistakes, or between different kinds of slips. With regard to the detection of different types of slips, Norman (1981) found that slips that result from failure to perform some action are more difficult to detect than slips that result from an inappropriately executed action. Norman (1981) also found that detection of a discrepancy between expectations and occurrences becomes more difficult if intentions are specified at a different, higher level than the erroneous action. According to Embrey and Lucas (1988), it is easier to recover from slips than from mistakes. In a study of computer users, Zapf et al. (1994) also found that errors on the intellectual level and recognition errors were difficult to detect, while errors at sensorimotor (skill-based) level were easy to detect. In a diary study involving everyday errors, Sellen (1994) noted that errors that were caught in the act most often involved slips, and that errors involving the wrong plan or intentions were more often caught on the basis of the outcome. Bagnara et al. (1988) found that in their study of recovery processes in a hot strip mill, slips were most often detected via external feedback and corrected without any causal analysis, while in recovery from rule-based mistakes at least some causal analysis took place, and for knowledge-based mistakes people either engaged in an exploratory causal analysis or they simply tried to overcome the resulting mismatch. In this thesis, I want to add to these insights, by exploring what the influence is of failure types other than human failures on the recovery process that follows.

Inherent to the failure or combination of failures that has led to the deviation or problem situation that needs to be recovered are the *potential* consequences if nothing is done about this situation. In risk analysis or risk assessment, risks are estimated based on the frequency or probability or likelihood of occurrence, and the severity, nature and significance of the expected outcomes (see for example Hoyos & Zimolong, 1988; Kumamoto & Henley, 1996, Kirwan, 1994; Cox & Tait, 1991). These expected outcomes correspond with the potential consequences if nothing is done, if no recovery steps are taken when they are needed.

In the literature, several outcome or consequence severity rating methods have been proposed, with fatalities among the general public often considered to be the most severe outcome possible (e.g. Kumamoto & Henley, 1996). Of course harm to people is not the only possible type of negative outcome. Outcomes can vary from injury or health effects (including death) to environmental damage, material damage, production or quality losses, and delays or time loss. Combinations of different types of outcomes are possible, too. Given a certain failure or combination of failures, even if no recovery steps are performed, several different outcomes may be possible, each with their own occurrence likelihood. Even without any attempt to recover, the worst case scenario does not necessarily have to take place: this may depend on factors such as circumstances and additional things that spontaneously happen in the rest of the process leading to the consequences, which can either deteriorate or improve the situation. The tighter coupled the different parts of a production system are and the more complex the interactions of these parts with each other, the greater is the potential of failures to lead to disastrous consequences (Perrow, 1984).

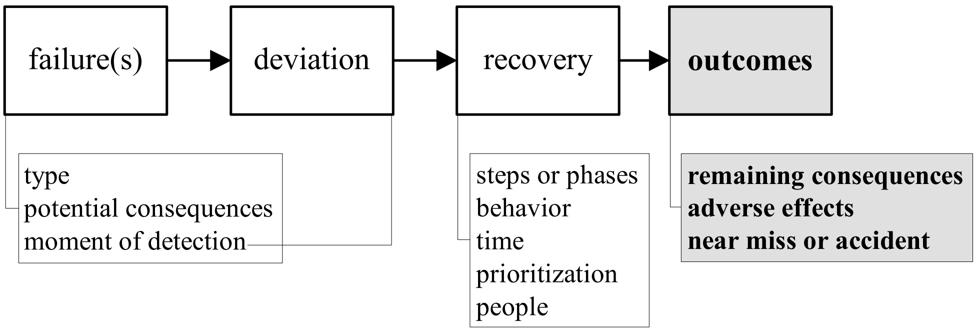
Similar to how this is done in formal risk assessments and safety investment planning, type, severity, and likelihood of potential outcomes may also play a role in what people decide to do when actually confronted with a situation where recovery is needed - as was pointed out already in subsection 2.2.4. In such situations however, depending on the time available before consequences will occur and in light of all other issues competing for attention, decisions may have to be made in a much more ad-hoc manner, and are specifically made for the situation at hand.

Another aspect of what happens leading up to the recovery process is the time elapsed before a deviation is detected. Recovery may become more difficult if more time has passed by since the occurrence of the failure(s), that is, since the deviation came about. If there is only a limited amount of time available before the deviation leads to negative consequences, there may not be enough time left to recover on time. Or, when it is necessary to know what caused the detected deviation in the first place before the appropriate corrective actions can be planned and implemented, this may become harder if more time has passed since. Waller (1999), for example, found a negative association between the amount of time taken to engage in adaptive responses to non-routine events and performance, in a study in which airline crews engaged in a high-workload flight simulation. Related to this plea for early detection is the performance stage in which errors are detected: both Sellen (1994), Kontogiannis (1999), Bove (2002), and Bove & Andersen (2001) distinguished three different stages of task performance during which errors associated with that task may be detected: during the planning stage, the action stage (while carrying out the task), or the outcome stage (when the results of the actions become noticeable). From a safety and reliability perspective, errors are best caught in the earliest performance stage possible, to avoid difficulties in the recovery process, as long as what happened concerned a real error, not something incorrectly perceived as such.

Summarizing what was said above: both the failures that have led to a deviation in need of recovery, their potential consequences, and the moment of detection may have implications for the recovery process that follows. To establish which relationships exist between all of these variables, I posed

> *Research question 6:*
> *How are*
> *a) the failure types that precede a recovery process in a working environment,*
> *b) their potential consequences,*
> *c) the moment of detection (how long the deviation already existed, and performance stage when errors are detected), and*
> *d) the form the recovery process that follows takes with regard to steps, behavior, and people involved, prioritization and time spent,*
> *interrelated?*

2.4 Outcomes of recovery processes



Not all recovery processes are always equally successful. In cases where all the necessary recovery steps are performed in a timely and effective manner, all the undesired negative consequences are avoided (recall the definition of a near miss given in chapter 1). Sometimes, however, it may not be possible to avoid some of the potential consequences of the failures from which recovery was needed. Even in the case of an accident, recovery actions may have been performed, which may at least have limited the consequences, so that the end result is less serious than it could have been. Recall that the term incident covers both near misses and accidents.

One way to measure how successful a recovery process has been, is to look at the outcomes of the process: what are the remaining consequences, if any, and were there any adverse effects of well-intended recovery steps? Very few researchers who have focused on recovery processes have actually included the outcomes of recovery process in their research. Bove (2002) and Bove and Andersen (2001) are exceptions. They distinguished inconsequential events (i.e. successfully recovered), undesired states, and additional errors as potential outcomes of recovery processes, depending on whether the people involved managed to trap the error, or not entirely, or actually made things worse.

To determine which factors influence the outcomes and thus the success of recovery processes, I formulated the following research question, combining all the aspects of recovery processes that have been discussed so far, including events leading up to them, to see how these relate to the process outcomes, i.e. the only new variable introduced here:

> *Research question 7:*
> *What is the relationship of*
> *a) the failure types that precede a recovery process in a working environment,*
> *b) their potential consequences,*
> *c) the moment of detection (how long the deviation already existed, and performance stage when errors are detected), and*
> *d) the form the recovery process that follows takes with regard to steps, behavior, and people involved, prioritization and time spent,*
> *with recovery process outcomes?*

The answer to this last question will bring all the insights gained from answering the first six research questions together and can be used as a basis for guidelines how to improve support for recovery processes through organizational and system design.

3 Initial studies[2]

3.1 Purpose and outline

To answer the seven specific research questions presented in chapter 2, initially a total of three empirical studies were performed. This chapter starts with a section in which the rationale and research design for these studies are described. The decision to use a multiple case study approach, which is addressed in subsection 3.2.1, was the first step in the research design. The next subsection, 3.2.2, contains a description of how organizations were selected for participation in the research. Subsection 3.2.3 addresses the data collection or sampling approaches used within these organizations. In subsection 3.2.4, the consequences of the choices that were made regarding data collection locations and methods are briefly discussed. After that, in subsection 3.2.5, an overview is given of all the variables included in the studies, along with a description of how the collected data were coded into the applicable values for each of these variables. Subsection 3.2.6 addresses how the collected data were analyzed. The findings that resulted from the analysis of the collected data, with regard to each of the specific research questions, are presented in section 3.3. A separate subsection is dedicated to every question. The findings are addressed in the same order as the order in which the research questions were presented in chapter 2, for the same reasons as given at the beginning of that chapter: first several aspects of the recovery process as main research topic itself (subsections 3.3.1 to 3.3.5); then the events preceding a recovery process and what their impact on the subsequent recovery process is (subsection 3.3.6); and finally the recovery process outcomes or end results and what the influences are of both the recovery process and the events that precede it on these outcomes (subsection 3.3.7). In section 3.4, these findings are interpreted and conclusions are drawn. This last section can also be read without reading section 3.3 first.

3.2 Research design

3.2.1 Multiple case study approach

As can be concluded from the previous chapter, even though some specific recovery process aspects had already received the attention of a few researchers, and many others had at least pointed out the potential benefits of focusing on recovery, at the start of the research described in this thesis there were still quite a few gaps in the existing knowledge

---

[2] Some of the findings presented in this chapter have also been addressed elsewhere: Kanse & Van der Schaaf (2001b; 2001c); Kanse, Van der Schaaf, & Rutte (2003).

about recovery processes. This meant that much of the research that needed to be performed to answer the research questions presented in the previous chapter would initially have an exploratory nature. As during that research more insights would become available, at a later stage the research could become more descriptive and even explanatory.

While the unit of analysis associated with most of the research questions was that of the recovery process, some of the questions also referred to lower level units of analysis: that of a single recovery process step or phase, and that of the separate failures preceding the recovery process. As was pointed out in the introduction, recovery processes occur both in work and everyday life settings. I focused my research specifically on recovery in work environments, since those processes do not only affect the person(s) involved in the process, but also the organization in which they take place. Organizations in their turn can use further insights into recovery to promote recovery possibities and thus improve organizational safety and reliability and even performance in general. Given this decision to study recovery in work environments, the organizations where recovery processes were studied formed an additional higher level unit of analysis.

A research method that lent itself perfectly to the situation described above was case study research (Yin, 1994), in particular the form in which a design is used whereby lower level units of analysis are embedded in higher level units. Since the main objective of the research described in this thesis was to draw conclusions with regard to recovery processes, not entire organizations, the main unit of analysis was the recovery process itself, not the organization where these processes took place, and every single recovery process was treated as a separate case. Multiple cases were included in the research to ensure that the resulting insights would be as complete as possible. Organizational characteristics were used, however, to explain differences in recovery processes between organizations, as can be seen from subsections 3.2.2, 3.2.4, and 3.4.2. The use of a case study approach also provided the opportunity to collect a wealth of data on single recovery processes, which was especially useful in the earlier stages of the research, when it was not quite clear which information was important and which not. In addition, the multiple case study approach allowed for fine-tuning and adaptations as the research progressed. Various data collection methods could be used simultaneously or sequentially. In fact, it is advisable to use different data collection methods and data sources in case studies, to increase construct validity (Yin, 1994), i.e. to establish correct operational measures for the concepts being studied. This approach is called triangulation (of data sources and data collection methods). Finally, using a multiple case study approach provided the possibility to study recovery processes in natural, real-life settings. Even though I could not control the events studied, at least there would not be any problems related to transferring laboratory findings to the real world – which is one of the key warnings from the domain of naturalistic decision making (e.g. Orasanu & Connolly, 1993).

As Yin (1994) pointed out, in multiple case studies, the logic that should be used to select cases differs from sampling logic commonly used in surveys, where respondents selected via statistical procedures are assumed to represent a larger population. Yin (1994) argued that replication logic should be used instead: cases should be selected so that they either predict similar results (literal replication) or contrasting results but for predictable reasons (theoretical replication). This replication logic helps to ensure the external validity of the research findings, that is, the generalizability of the findings beyond the cases actually studied. Similarly, Eisenhardt (1989) argued that theoretical instead of statistical sampling should be used for theory building based on case research. Based on this replication logic, in my research, at first I aimed at mostly literal replications, in the same environment and circumstances, and at a later stage I also aimed at theoretical replications, by specifically including a wider variety of cases, and cases that took place under different circumstances, or even in a different environment, that is, in a different organization. This approach did not only enable a comparison of recovery processes within one organization, but also between organizations.

3.2.2 Selecting organizations

High-risk, high reliability industry

To optimize the chances of finding information about recovery processes, I decided to specifically look at recovery processes occurring in high reliability organizations (for a description, see Roberts & Bea, 2001; or Weick, Sutcliffe & Obstfeld, 1999). The accident frequency in high reliability organizations is so low (hence the name) that the accidents that do take place do not generate sufficient learning opportunities for improving the organization's safety and reliability. To capture additional learning opportunities, in the last decades many high reliability organizations have set up a system in which near misses (where negative consequences were avoided) are reported and analyzed (see, for example, Van der Schaaf et al., 1991, page 2). So far, high reliability organizations have primarily and successfully used near miss reporting or similar systems to identify errors and other failures against which they need to implement preventive measures (e.g. Van der Schaaf et al., 1991). A much less known and barely explored benefit of near miss reporting systems is that they have the potential to provide insight into the reasons why the recorded near misses did not develop into accidents, into the recovery steps (if any) that were taken, and into the factors that influenced the recovery process. To my knowledge, Bove (2002) and Bove & Andersen (2001) are among the first researchers who have used near miss data from actual work environments collected via already existing reporting systems to analyze recovery in the domain of air traffic control.

One could believe that the preventive measures high reliability organizations have implemented against identified risks are the only reason for their achievements in the area of safety and reliability. But still, problems can and will occur in such organizations. Another part of the reasons for the low accident frequency of high reliability organizations may lie in their ability to recover from problems once they occur; which would make recovery an important additional strategy in reaching their high levels of reliability. This idea is supported by the parallels between one of the characteristics of a high reliability organization as described by Roberts and Bea (2001) and the work on error management and error training by, among others, Frese (1991), Van Dyck (2000), and Dormann & Frese (1994). These error management researchers argue that the error management culture of an organization, that is how errors are viewed and dealt with in that organization, plays an important role in their ability to recover. Their definition of a positive error management culture (associated with a great ability to recover from errors and resulting problems) is a culture in which errors are openly discussed and viewed as learning opportunities, rather than negative experiences to be ashamed of. Roberts and Bea (2001) observed something very similar in the way high reliability organizations deal with the possibility of problems in their systems: these organizations realize the need to involve their people in the never-ending search for both actual and potential problems. They prepare and empower their people to recognize, communicate, and deal with problems. According to the theory on error management culture, precisely this characteristic typifies organizations that are good at recovery.

Based on everything mentioned above about high reliability organizations, I concluded that in such organizations I could not only be sure that recovery processes actually occur, but also expect some information on such processes to be available in the form of near miss reports. While most of the definitions provided in the literature for high reliability organizations included only organizations with high-risk processes, where failures can rapidly lead to devastating consequences (e.g. Weick et al, 1999), some definitions were a little more flexible with regard to which organizations still qualify, and included those with less risky operations as well. For the purpose of my research, I expected the benefits of recovery processes to be most visible in organizations with high-risk processes, such as nuclear power plants, chemical industry, aviation and aerospace. From such types of industries, access to the chemical industry turned out to be easiest.

Study 1

The first organization to participate in my research was a large chemical process plant in the Netherlands. This plant had roughly 200 employees. It started operations in 1972 and had expanded ever since. Its main products were plastics, synthetics and foams used in packaging materials, coatings, furniture, cosmetics, automotive parts et cetera, which were produced in a continuous process. All production personnel worked in teams, in continuous operation, with eight hour shifts. Other personnel at operational level, such as maintenance staff, and electrical and instrumentation engineers also worked in teams. Within these teams, most tasks and responsibilities were shared between several people. Asking for (and offering) assistance, both inside and outside one's own team, was a commonly accepted and regularly used practice. Procedures were available for how to respond to most of the known, possible deviations. The plant had an excellent safety record. A voluntary near miss reporting system had successfully been established there and had been operational for seven years. Via this system, production plant and other personnel as well as contractors reported near misses and minor incidents in which they were involved. Accidents, the more serious events with regard to negative consequences that occurred, were reported and followed up via different channels. In the near miss reporting system, reporting forms were used on which time, date, and location of the event, and a few lines containing a description of what happened, how this could happen, potential consequences, steps taken, and recommendations were recorded. The reports were entered into a computer-based near miss and minor incident database by safety department members. This database was also used by the safety department to identify areas in need of attention, based on the analysis of factors that played a role in larger sets of near misses and minor incidents, combined over a longer period of time. Other examples of the use of this database included the assessment of the effects of implemented safety measures. The steady flow of near misses reported in the system, on average two per day, formed a good indication of its success. Key factors for the success of the system were the 'no blame' culture of this company regarding reported errors, the anonymity of the reports once analyzed for their root causes and entered into the database, and the quick follow-up by the appropriate personnel in response to reports. I used this near miss reporting system as a basis for my first study.

Study 2

Even though the cases collected in the first study contained a wide variety of recovery processes, involving recovery from various types and combinations of failures, also human error, there were no cases among them involving recovery from specifically *self-made* human errors. By recovery from self-made errors I mean recovery from errors made by the same people as those involved in the subsequent recovery process, as opposed to recovery from somebody else's errors. The plastics plant's safety department confirmed that this was representative of all the reports in their near miss database. Most of the existing research into recovery processes, however, as was pointed out already in section 2.3, had focused exclusively on recovery from one's own errors. A comparison was complicated even further by the fact that different variables were used in these studies. To enable a comparison between recovery from self-made errors and from other (combinations of) failure types, and to gain a more complete picture of the recovery process, I set up a second study in the same plant that had already participated in the first study. This study started roughly one year after the end of the first study.

Study 3

To allow a comparison of recovery processes between different organizations, a year after the second study I performed a third study, in another chemical process plant, this time in Belgium. This plant was an oil refinery that had circa 90 employees, including production-, product handling-, maintenance-, research and development-, quality control-, and sales personnel. Production and product handling personnel worked in teams, in continuous operation, with 12 hour day and night shifts. The teams of mechanical engineers, electrical engineers, and instrumentation engineers worked daytime, and lab technicians responsible for quality control and research and development worked in morning and afternoon shifts. The teams in this plant were smaller than those in the plant where the first two studies were performed. Production team members all had their own specific tasks, which they would share with one other coworker on days without absenteeism, while on other occasions the task could be solely their own responsibility. Asking for assistance, both inside and outside one's own team, was an accepted practice, but in this plant there were less people to approach in such cases than in the plant where the first two studies were performed. The refinery specialized in the production of bitumen. Bitumen is one of the heavier fractions of crude oil and one of its main applications is in asphalt for road surfaces. Given the high temperatures required to keep bitumen in a liquid state and the associated long startup times for the installation, the refinery produced in a continuous process. Special blends were made in a separate part of the plant in batch processes. While most of the plant's supplies arrived via ships, the final products were transported to customers via trucks, so both unloading and loading operations took part on-site. Most of the risks at this plant were linked to the high production process temperatures: for example, when hot bitumen comes into contact with water it will start to boil and increase to ten times its original volume, and when bitumen gets cold it will block the pipelines through which it flows. Procedures were available for how to respond to most of the known, possible deviations. Even though this plant did not have a system specifically aimed at the reporting of near misses, sometimes near misses were reported via the so-called ‘continuous improvement system’ that had been set up by the quality department. The reported near misses most often involved either the more visible events, or situations where a very clear need for improvements existed and the person reporting the situation wanted to focus the organization's attention on it. In this plant, it was much less common to discuss failures, especially those for which people might be blamed, with management. Among immediate co-workers, however, the attitudes towards discussion of such topics were much more open than when managers were involved in the discussion.

3.2.3 Data collection methods

Study 1

In my first study, I used the reports contributed to the existing near miss reporting system of the Dutch plastics plant described in subsection 3.2.2 as the basis for the data collection process. Starting with the near misses that were reported on the first day of data collection, and after introducing my research project to the plant's personnel, I set out to collect data during the days that followed on as many different recovery processes as possible, with different types and numbers of steps involved, different behavior levels, different amounts of time spent on the separate steps and the entire process, different numbers of people with different organizational roles, different preceding failures and potential consequences, different detection moments, and different outcomes - all of this to find out which similarities existed within such a variety of processes and what the effects were of the differences. Follow-up interviews were held with those involved in the reported near misses to obtain additional information. Occasionally I spoke to more than one person for a given report. This could occur, for instance, when multiple people were involved in the recovery steps. I analyzed a total of 50 cases. These were literally all the near misses reported by the different teams of operators, other personnel and contractors, within a 24-day time frame. At that point I had reached a situation where newly reported recovery processes began to resemble earlier reports, that is, no values were found any more for any of the variables used to describe the recovery process (see subsection 3.2.5) that had not been found several times before. As it turned out, all the cases involved recovery from actual deviations. There were no examples of cases where, in retrospect, people realized that a situation was incorrectly perceived as a deviation and where unneeded recovery steps were implemented. I always held the follow-up interviews needed for further information on the recovery process within two weeks after an event was reported. During my study, unusual events such as plant shut downs were neither scheduled, nor actually occurred. A complete list of the topics on which I collected information via the plant's report forms and follow-up interviews is given in subsection 3.2.5, where the variables included in all of the three initial studies are presented.

Study 2

Since the plastics plant's near miss reports were not likely to provide insight into recovery from self-made errors, even if I would wait for more events to be reported to the system, I used a different data collection approach in the second study. Inspired by Sellen (1994), who studied detection of everyday errors using diaries, Reason & Lucas (1984) and Reason & Mycielska (1982), who also used diaries in their error studies, for this study I used event diaries. I asked one team of operators from the same plastics plant that had already participated in my initial study, including operators from all the different sections of the plant, to keep track of how they, more or less successfully, recovered from their own errors. To construct the diaries, I specifically designed proformas to collect the same information about the error and subsequent recovery steps as I collected via different methods in the initial study. As was pointed out before, a complete list of these topics is given in subsection 3.2.5. An example of such a proforma is included in appendix 2, translated from Dutch into English. For the errors they made and recovered from, the operators were asked to fill out a proforma in their diary, during a spare moment as soon as possible after the completion of that recovery process. I ensured them that only I would read the cases they reported, for research purposes, that I would be the only one who could trace back diaries to persons, and that their managers would only receive feedback over the dataset as a whole, not about single cases. I asked them to participate in this study for a period of 15 working days (5 afternoon shifts, 5 night shifts, and 5 morning shifts). During this time normal plant operations were expected, special events were neither scheduled nor actually occurred. In total, 21 participants reported 33 events involving recovery from self-made errors via the diary study. Again, all of these events involved recovery from actual deviations, not from situations about which people in retrospect realized that they were incorrectly perceived as deviations. Most likely, in this organization the latter type of events did not occur very often, after all, in the first study no examples were found either. On the other hand, since I had no specific research questions that focused on unnecessary recovery, I never explicitly told the participants whether I wanted information about such cases or not.

Study 3

To collect data on recovery processes in the Belgian bitumen plant, I used interviews based on the critical incident technique (Flanagan, 1954). In these (confidential) interviews, I told people about my research objectives and asked them to tell me about an event they could still clearly remember in which they had successfully (or at least to some extent) recovered from a situation that would otherwise have led to (more) serious negative consequences. Some of the interviews were prompted by reports in the plant's continuous improvement database. For the topics that were addressed in the interviews, again, refer to subsection 3.2.5. I collected data on a total of 39 cases involving recovery from actual deviations. In this study, too, the collected data did not include cases involving unnecessary recovery from situations about which people in retrospect realized that they were incorrectly perceived as deviations. Such events were most likely quite rare in this organization as well. Just as in the other organization, I never explicitly told the people I interviewed whether I wanted information about such events or not.

3.2.4 Expected effects of chosen data collection locations and methods

Both the decision to collect data in specific organizations and the specific data collection methods that were used in the different studies had some consequences for the types of cases involving recovery that were likely to be collected in each of the studies.

As for the influence of organizational characteristics, for example, the culture with regard to discussing and reporting failures, even those that were successfully recovered, was expected to have a substantial effect on the collected types of cases. This culture seemed to be more open in the plant where the first two studies were performed than in the plant where the third study was performed, so in the former plant I expected to have less difficulties in finding data about all kinds of failures, even less visible ones and ones with only minor potential consequences, than in the latter plant. Also the accepted practice with regard to asking for and offering assistance when dealing with deviations, and the availability of such assistance were expected to be influential, especially with regard to who and how many people were involved in the recovery processes, and how long the processes took. Furthermore, the availability of known ways to respond to deviations was obviously expected to influence the amount of pre-planned behavior in recovery steps.

As for the effect of the data collection method used on the collected types of cases: in the first study, cases were sampled based on their occurrence within a certain period of time, that continued until newly added cases did not substantially increase the variety in the dataset any longer, with the restriction that only those cases that were also reported to the existing reporting system ended up in the dataset. If there would be a certain threshold to reporting, this would affect the collected types of cases. Another possible approach that would still involve the use of this reporting system, could have been to sample on the basis of types of failure involved. Such an approach, however, could easily lead either to having to wait a long time or having to trace back far in time, if a certain type of failure does not occur very often, with the associated delay, or difficulties to recall the event. By choosing a fixed time period such problems would not play a role, and the resulting dataset would much better reflect the types of cases involving recovery that normally occur within the organization. The effect of the data collection method used on the collected types of cases was most obvious for study 2. By specifically asking for errors the same person(s) both made and recovered, no cases were expected involving recovery from other types and combinations of failures. Possibly the cases collected in this study would also have a less serious nature, otherwise they might normally be reported already via the existing reporting system in this organization, which hardly contained any cases of this type. In the third study, a much less structured approach was used to sample cases. I expected a certain amount of coincidence to play a role, because I never knew before I started an interview whether the person I was going to interview had been involved in a recovery process recently or not. Nevertheless I made an effort to talk to all personnel, to collect as much data as possible. Since the people in this plant were less familiar with the concept of near misses and recovery than in the plant where the first two studies were performed, I expected that most of the collected cases would involve more serious types of events, which are generally also easier to recall.

In each of the subsections of section 3.2, where the findings are presented for each of the specific research questions formulated in chapter 2, where applicable, some attention is also given to the identified differences between the three studies with regard to those findings. In subsection 3.3.2, these differences are interpreted and explained in terms of how they could have been the result of either organizational characteristics of the plant where the data were collected or the data collection method used.

3.2.5 Variables and data coding

Even though in each of the three studies a more or less different method was used to collect data, through the questions that were asked in each approach I ensured that each of these methods produced the same information about recovery events. As a result, to answer the detailed research questions that were presented in chapter 2, I could use data from all three studies. Before starting the coding of the data, in each of the three studies I had my informants check the case reports I made based on their information, either by reading my notes back to them or by having them proofread the documented case.

Variables

To enable a structured comparison of the recovery processes about which I had collected information during my studies, I coded the data for each case or event into a number of variables. During data collection, I tried to get as much information as possible with regard to these variables. The variables relevant for the research questions presented in chapter 2 were:

- Recovery related actions and cognitive processes were considered as a separate process step if they had a different goal than the immediately preceding recovery efforts, so that based on this goal they belonged to a different phase. They were also considered as a separate process step if they were performed by different persons than in the previous step, or at a different moment in time, for example when an unrelated action was performed in between, or when at least for a while no efforts were spent on recovery. Note that in every case or event multiple recovery steps were performed – every separate recovery step can also be treated as a lower-level unit of analysis. For every separate recovery step that was performed during the recovery process and that could be distinguished based on information obtained from the people involved, the following set of variables were included:

- o *goal of the process step* (detection, explanation, or countermeasure(s));
- o *whether the step involved (pre-)planned, unplanned (ad-hoc), or a combination of both types behavior*, from the point of view of the organization (see description of these types of behavior in subsection 2.2.2);
- o and *whether the step involved human information processing behavior at skill-based, rule-based, or knowledge-based level* (as defined by Rasmussen (1976, 1986, and described in subsection 2.2.2); in case of doubt always the highest likely level was chosen, i.e. the level involving the highest applicable amount of conscious information processing).
- o the *number of people involved* in that particular step;
- o their *organizational roles*;
- o whether they were the s*ame, some of them the same, or different people as/than those involved in the previous step*;
- o for steps in which others became involved the *reason for involving these others* (free text that was afterwards divided over categories based on both the literature and what subsequently emerged from the data);
- o the amount of *time spent on the process step* (categories were used, cf. total amount of time).

- The *order in which the applicable types of process steps or phases involved in the recovery process occurred*.
- The *total number of steps* taken during the recovery process.
- *Total amount of time between start and completion of the recovery process.* I used categories of time periods, starting with up to 5 minutes (i.e. almost instantly), 5 minutes to 1 hour, 1 to 8 hours (duration of one shift), 8 to 16 hours (next shift), 16 to 24 hours, 1 to 7 days, 1 to 4 weeks, and as a last category, longer than 4 weeks. These time categories were based on what the people involved were able to remember. Generally they could not recall the exact amount of time, but they were able to indicate a global category in which this amount would fit, that is, was it a matter of minutes, hours, days, weeks, months, et cetera, the commonly used measures of time, and was it within one's own shift, the next one, or even later. The border between the

first and second category, five minutes, was used to distinguish between quick steps or processes and those that obviously required more effort. Most often, people were not sure whether they were ready within a minute or just outside that time span, which is why this border was set at five minutes. The longest time categories were included based on the understanding that even though separate process steps might not take very long by themselves, because of time that might elapse in between efforts that actually aimed at recovery, for example as a result of other tasks or having to wait for new parts to arrive, the entire process might take a very long time before it is completed.

- *Amount of tasks other than recovery requiring attention* (no, some, or many other tasks also require immediate attention).
- The *types of failures* causing the deviation or dangerous situation from which recovery was needed. In line with what was said in section 2.3, I made a distinction between technical, organizational, and human failures. Within the category of human failures, I distinguished skill-, rule-, and knowledge-based failures. I also included a rest category of other (including external) failures. In the first two studies, I was assisted in the identification of the failure types involved by the plant's safety department staff, who had an in-depth understanding of the causal factors involved in the near misses, as a result of their many years of experience in in-depth analysis of the plant's near misses and accidents. In the third study I was assisted by the plant's training coordinator who had many years of operational experience and an in-depth knowledge of all the possible tasks in the plant and how things can go wrong. For each main category of failure types, I recorded the presence or absence of failures of that type. Failure types involved in a near miss or minor incident were categorized from the point of view of the person or people involved in the event. As an example, though a pump failing may ultimately be the result of a human error by the designer, as the designer is not directly involved in the resulting problem and the design process is outside the control of the operators confronted with the failure, the failure is categorized as a technical failure. Note that more failure types are possible within a case or event (and can be used as a lower-level unit of analysis).
- The *time elapsed between failure occurrence and detection*. I used the same time categories as mentioned for total amount of time.
- The *task performance stage* during which the failure was detected: planning, action, or outcome stage.
- *Time until the earliest possible occurrence of potential consequences* if the failures were not to be recovered. I used the same time categories as for total amount of time, plus one for situations where it was impossible to say when this moment would be. This could be the case when the actual occurrence of the consequences depends on other factors than simply time elapsing, such as in the situation where a piece of

equipment which is only needed during certain operations, is found to be broken: whether this leads to negative consequences or not depends on it actually being needed while it is still broken.

- *Possible consequences* in the absence of recovery: i.e. *type* of consequence (production/quality loss, delay, damage, injury/health effects, environmental spill) and a *severity* indication for each type (no (0), minor (1), considerable (2), or major (3)). (The likelihood of these consequences actually occurring turned out to be too difficult to establish for the people involved, whom I interviewed afterwards and asked about this probability. Possible consequences were simply defined as those consequences that *can* occur based on the detected deviation.)
- *Remaining consequences* after recovery steps: as above, plus a rough indication of costs for repair, and whether the hazards involved continued to exist for a while or not.
- *Unintended adverse effects* of recovery steps: presence and severity (no/minor/considerable/severe adverse effects).

Coding process

The first stage in the coding process was the development of a set of coding instructions for all of the variables listed above. These instructions were compiled by coding a small number of cases together with another researcher, an expert in human reliability and incident analysis. For this exercise, the cases were selected to reflect the variety of types of recovery processes, to ensure that every possible lack of clarity with regard to coding was discovered and dealt with in a way acceptable for both coders, before they continued with the coding of other cases separate from each other, i.e. not in each other's presence. Interrater reliability checks were performed on subsets of the data collected in all three studies.

For the first study, I coded all 50 cases and the other coder, who had also assisted in development of the coding instructions, coded 25 % (13) of the cases, other than the cases used for the development of the coding instructions. For these 13 cases, all of the variables were coded by both coders. This subset was a representative sample, reflecting the variety of process types among the total number of cases. Coding the data was a time-consuming process and if sufficient agreement could be reached for a sample this size, it would be unlikely to find disagreement for the total dataset. Where differences in coding were identified, consensus was reached between the coders about the final coding to be used for further analysis of the data. To measure the agreement between the two coders (before the consensus was reached), I calculated Cohen's kappa (Cohen, 1960), a statistic in which agreement levels are corrected for chance agreement, for each variable separately. Landis & Koch (1977) suggested that values of the kappa statistic between 0.41 and 0.60 indicate moderate agreement, between 0.61 and 0.80 substantial agreement, and above 0.80 almost perfect agreement (1 reflects perfect agreement between raters). We found almost perfect agreement for 12.5% of the variables (the smallest kappa value was 0.822, $p<0.001$), and perfect agreement for 87.5% (Kappa values of 1, $p<0.001$).

Based on the distribution of the failures involved in the 50 cases over the categories technical, organizational, human and external/other failures, another check was performed on the representativeness of the 50 selected cases with respect to the plant's near miss database statistics. The number of failures in each category was 33, 44, 35, and 4, respectively, which was, according to the plant's safety department staff, a distribution quite common for this process plant. Note that more than one failure type can be attributed to one case.

For the second study, similar to how this was done in the first study and based on the same arguments, an interrater reliability check was performed on a representative sample, 9 (25%) of the 33 cases, by calculating kappa values for each variable separately. Again, where differences in coding were identified consensus was reached between the coders about the final coding to be used for further analysis of the data. For 61.4% of the variables we found perfect agreement (kappa values of 1, $p <0.001$) between the two coders. For 11.4% of the variables we found kappa values in the 'almost perfect agreement' range, for 22.7% of the variables values in the 'substantial agreement' range, and for 2.3% of the variables values in the 'moderate agreement' range (p between <0.001 and 0.047). For only 2.3% of the variables initially no agreement was found between the two coders. The lower kappa values compared to the first study were the result of less detailed data being available per case. In the interviews there was an opportunity to ask for more information, whereas the amount of detail in the diaries depended on what the participants decided to write down.

In the third study, an interrater reliability check was performed on 28 (72 %) of the total of 39 cases but only with regard to the goal of each of the separate recovery steps of every case. Those were the key variables in my research, and the first two studies had shown that most of the other variables were not often subject to disagreement between coders, especially the more objective variables like people involved, and time-related issues. Every recovery step taken in each event (a total of 123 steps for 28 events) was coded both by myself and a research assistant, using the set of definitions or coding instructions developed during the first study to assign a step to a specific category based on its goal. The inter-rater agreement between us, expressed as a kappa value, was 0.92 (significant at $p<0.05$), signifying almost perfect agreement. For those steps where we differed, consensus was reached regarding the category before further analyses were performed.

3.2.6 Data analysis

The nature of the variables included in the research determined which methods could be used to analyze the data. As can be seen from their descriptions in subsection 3.2.5, most of the variables that were included in the initial studies had a categorical character: the goal of the recovery process step; the type of behavior involved (pre-planned, ad-hoc, or a combination; and skill-based, rule-based, or knowledge-based); the organizational roles of people involved; the reason for involving others; the order in which the process steps were performed; the preceding failure types; the types of possible consequences; and the types of remaining consequences. The majority of the other variables had an ordinal character: whether the people involved in a particular process step were the same, some of them the same, or different people then in the previous step; the amount of time spent on a recovery step; the amount of time between start and completion of the recovery process; the extent to which other tasks also required attention; the amount of time elapsed between failure occurrence and detection; the task performance stage during which the failure was detected; the amount of time until the earliest possible occurrence of potential consequences; the severity of each of the distinguished types of potential consequences; and the severity of each of the distinguished types of remaining consequences. Only two of the variables were ratio scale variables: the number of people involved, and the total number of steps taken in a recovery process.

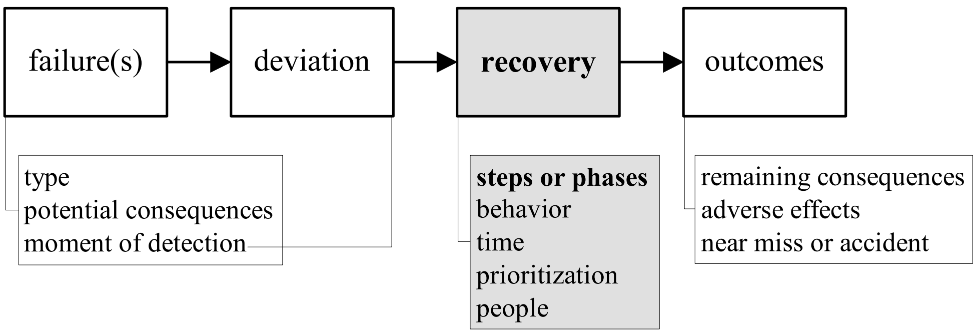
To establish relationships between two categorical variables, chi-square tests were used. Where necessary to draw further conclusions, the adjusted standardized residuals were analyzed, derived from the differences between observed and expected counts in the cells of the cross-table of the two variables, to identify where the distributions across the cells differed from what was expected. For this analysis, the expected cell counts were calculated based on the table's row and column totals, the marginal totals. To establish relationships between categorical and ordinal variables, nonparametric test procedures were used, as the assumption that the ordinal variables had a normal distribution, necessary for parametric tests, was not met. Kruskal-Wallis tests were used as overall or omnibus tests, whereby a chi-square statistic was calculated to evaluate if the mean rank scores on the ordinal variable were equal for different values or categories of the categorical variable, and Mann-Whitney follow-up tests were used, whereby *z*-scores were used to evaluate pairwise differences between different values of the categorical variables with regard to scores on the ordinal variable. To establish relationships between two ordinal variables, Spearman correlations were calculated. I used Pearson correlations to establish relationships between ratio scale variables.

For most of the analyzed relationships between pairs of variables, the distribution of the cases across the corresponding cross-tables did not meet the assumptions necessary for the use of the asymptotic method to calculate the significance level, that is, for most cross-tables the percentage of cells with expected frequencies less than 5 exceeded 20%. Therefore, in these analyses, Monte Carlo estimations for the exact significance were used instead. For calculating estimations of exact significances no assumptions have to be met regarding expected cell counts. Unless stated differently, all significance levels presented in section 3.2 are estimations of exact significances.

In general, a significance level of α=0.05 was used for the analyses. However, for those Mann-Whitney follow-up tests that involved a relatively large number of pairwise comparisons, which increased the overall chance of finding at least a few significant differences, this level was reduced to avoid type I error, i.e. to avoid calling differences significant that might not really be significant. To this end, a slightly more pragmatical approach than the rather restrictive Bonferroni method to avoid type I error, or the less restrictive but laborious Holm's sequential Bonferroni method (e.g. Green, Salkind, & Akey, 2000) was used. For Mann-Whitney follow-up tests involving a substantial number of pairwise comparisons, i.e. where the categorical variable involved had more than three different values, the significance level used to evaluate each of the comparisons was set to α=0.01. Where this was the case, this is mentioned in the text.

3.3 Findings with regard to each of the research questions

3.3.1 Steps or phases involved in recovery: the process phase model



Recall that the first research question aimed at verification and further development of the recovery process model presented in figure 4 in subsection 2.2.1 and repeated below in figure 5: *a) With regard to the steps and transitions involved, does the proposed recovery process phase model accurately describe the processes people follow to recover from* *failures in work situations, b) which different types of recovery processes can be distinguished based on which phases occur when during the process, and c) which additional details, in the form of subgoals, can be distinguished for each of the phases of the proposed phase model?*

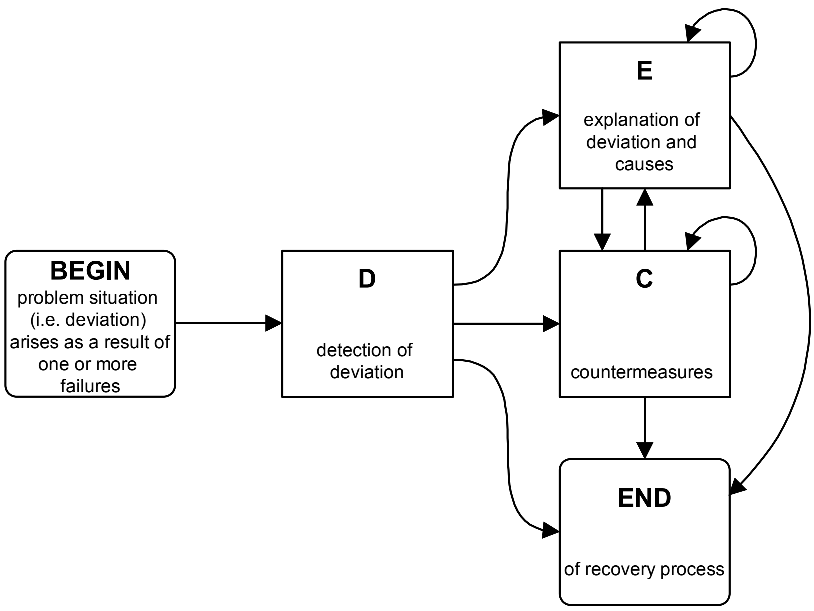


Figure 5 Proposed recovery process phase model

Model verification and categorization of recovery process types
To verify the existence of the proposed process phases, detection, explanation, and countermeasures, all the steps involved in the recovery processes about which information was collected in each of the three studies were categorized based on their goals as belonging to one of these three phases. At the end of the coding process there were no steps left that could not be assigned to one of these categories. In fact, assignment of the steps to one of the three phases turned out to be a very straightforward process, where never any doubt existed with regard to the choice for a category. This confirmed the accuracy of the model with regard to the three phases distinguished.

As for the proposed transitions between phases: the actual forms the analyzed recovery processes took differed widely between events. There were cases where after detection, only countermeasures were taken. An example of such a simple recovery process form was the case where a chemical process operator, arriving at the laboratory to drop off product samples for quality control, detected that he had forgotten to take a sample of product X (detection phase), after which he went back to the sampling point for product X in the plant, took the sample, and returned to the lab to deliver the product X sample (countermeasures phase). But there were also many cases that took a much more complex form. For example the case where a defect in one of the electronic circuits of a signal transmitter caused a product flow indicator to fail and disappear from the process control panel, which the control panel operator noticed (detection phase). To avoid an automatic shutdown of part of the plant and thus production loss, the control panel operator had to override the automatic shutdown setting on the flow level (countermeasure to stabilize the situation). To ensure that meanwhile the product flow would not cause any problems, the operator had to look for alternative indications in the panel to make inferences regarding the flow rate. The operator had to call in the instrumentation engineers to find out what caused the disappearance of the indicator. These engineers, once they had identified (explanation phase) the defect circuit board, looked for a replacement (countermeasures phase). They did not have any spares of the right type, so they used the circuit board of another transmitter from a part of the installation currently not in use due to maintenance (temporary countermeasure). They only thought of this solution after having consulted one of their co-workers with regard to the situation, who remembered having seen the required type of circuit board a few days earlier while he was involved in the maintenance activities in the shut down section. A new circuit board was ordered and installed a few days later in the transmitter from which it was borrowed (countermeasures phase), just in time before the scheduled start-up of that section of the installation.

Among the collected cases there were many examples of the simple form of a recovery process, that is, detection followed by one or several steps belonging to the countermeasures phase. There were even more examples of the more complicated types of processes, involving both efforts aimed at explanation and countermeasures after the detection phase, sometimes even including iterations. There were also some examples of cases where after detection, no further steps were taken, where further steps were not deemed necessary or appropriate, but no examples where the process stopped after detection and explanation. Some participants in the studies, however, confirmed that the latter forms of recovery processes also existed in their organization, that is, when during the explanation phase the decision was made that no further steps would be taken, because they were deemed unnecessary, not (cost-)effective, or impossible.

Based on the distinction made above between different forms or types of recovery processes depending on the occurrence or absence of certain recovery process phases, the recovery processes could be divided into three categories as follows:

- processes with only detection and no further recovery steps (D);
- processes with, after detection, only countermeasures ($D \rightarrow n*C$);
- processes that include, after detection, both steps aimed at explanation of the deviation and/or its causes, and countermeasures ($D \rightarrow \{k*E \ \& \ n*C\}$);

whereby n and k represent the number of countermeasures steps and the number of explanation steps, respectively. They are both integer numbers with as value 1 or larger, that is, n=1, 2, 3, ... and k=1, 2, 3, ... Bear in mind that not only steps with a different goal , i.e. belonging to a different phase, were treated as separate steps, but also steps performed at different moments in time or by a different person or group of persons. In addition, within the third category of processes involving both steps aimed at explanation and countermeasures after detection, two sub-categories were distinguished, based on the moment when the first step aimed at explanation took place: this could either be immediately after detection, or after some initial countermeasures were taken first, respectively:

- D→E→{(k-1)*E & n*C}
- D→C→{k*E & (n-1)*C}

In table 3, for each of the three studies the distribution of the collected cases over these four categories is presented.

Table 3 Number (and percentage) of cases per recovery process type, for each of the studies

	number (percentage) of cases					
recovery process type	**study 1**		**study 2**		**study 3**	
D	1	(2%)	1	(3.0%)	0	(0%)
D→n*C	12	(24%)	15	(45.5%)	17	(44%)
D→E→{(k-1)*E & n*C}	31	(62%)	15	(45.5%)	13	(33%)
D→C→{k*E & (n-1)*C}	6	(12%)	2	(6%)	9	(23%)
total	*50*	*(100%)*	*33*	*(100%)*	*39*	*(100%)*

This table shows that all of the transitions between phases represented by the arrows in figure 5 were indeed found within the collected cases, and, furthermore, that no other, unmodeled transitions were identified. A closer inspection of the parts {(k-1)*E & n*C} and {k*E & (n-1)*C} revealed some additional insights, as described below. Again, steps aimed at explanation proved useful in distinguishing between processes.

In study 1, in 20 out of 50 cases (40%) a second instance of the explanation phase took place (e.g. first by the operators, then by engineers) and in one of these cases even a third instance, whereas in study 2 a second explanation step only happened in five out of 33 cases (15%). In study 3, in seven out of 39 cases (21%) a second instance of the explanation phase took place, and in two of these cases even a third instance. Generally, that is in 20 of 32 cases involving multiple explanation steps, between two steps aimed at explanation some form of countermeasures was taken. Most often in those cases, the first explanation step had identified what needed to be done within a short time, before there would be time for a more in-depth investigation. The exceptions were four events in study 1, four in study 2, and four in study 3, where the explanation steps followed immediately after each other. Most processes ended with a countermeasures phase, except for the processes where only detection took place, and in study 1 another five processes of the $D \rightarrow \{k*E \; \& \; n*C\}$ type. In those five processes, the countermeasures could be completed relatively quickly. And after these countermeasures, some additional time was taken for a causal analysis of the event, to identify which lessons the organization could learn from what had happened. One example was found of a recovery process where the initial explanation phase that followed the detection of a valve not working as expected, made the people involved set off in the wrong direction. They thought they had found the cause in a loose wire. Only when fixing that did not work, they decided to call in an engineer, and just after making that call they discovered that a switch in the control room had been left in the wrong position, causing the whole problem. This they could correct easily, and the engineer could be called off.

Not only the types of steps involved and the order in which they are performed can be used to categorize recovery processes, but also the number of steps involved in the entire recovery process. Table 4 contains an overview for each of the three studies of the number of recovery process steps involved in each case, and in table 5 for each study the average number of steps per recovery process type is presented.

Table 4 Number (and percentage) of cases per total number of steps involved in the recovery processes, for each of the studies

	number (percentage) of cases					
total number of steps	**study 1**		**study 2**		**study 3**	
1	1	(2%)	1	(3.0%)	0	(0%)
2	1	(2%)	6	(18.2%)	6	(15.4%)
3	7	(14%)	12	(36.4%)	8	(20.5%)
4	12	(24%)	9	(27.3%)	8	(20.5%)
5	17	(34%)	3	(9.1%)	6	(15.4%)
6	7	(14%)	1	(3.0%)	6	(15.4%)
7	1	(2%)	1	(3.0%)	4	(10.3%)
8	4	(8%)	0	(0%)	0	(0%)
9	0	(0%)	0	(0%)	1	(2.6%)
total	*50*	*(100%)*	*33*	*(100%)*	*39*	*(100%)*

Table 5 Number of cases and average number of steps per recovery process type, for each of the studies

	study 1		study 2		study 3	
recovery process type	**# of cases**	**av. # of steps**	**# of cases**	**av. # of steps**	**# of cases**	**av. # of steps**
D	1	1	1	1	0	-
D→n*C	12	3.8	15	2.9	17	3.1
D→E→{(k-1)*E & n*C}	31	5.3	15	3.9	13	5.1
D→C→{k*E & (n-1)*C}	6	4.8	2	5	9	5.9
total	*50*	*4.8*	*33*	*3.4*	*39*	*4.4*

A Kruskal-Wallis test (χ^2 (2, N=122)=16.035, p<0.001) indicated that there were significant differences in the number of recovery steps between the studies. Follow-up Mann-Whitney tests revealed that there were significant differences between studies 1 and 2 and between studies 2 and 3, respectively z=-4.777, p<0.001 and z=-2.327, p=0.020, the mean ranks for study 2 were significantly lower than for the other two studies. These findings, not surprisingly, indicated that recovery from only self-made errors, the only types of event included in study 2, on average took fewer steps than recovery from other (combinations of) failures.

Another Kruskal-Wallis test (χ^2 (3, N=122)=44.815, p<0.001) indicated that there were significant differences in total number of steps involved between the four types of recovery process that were distinguished earlier in this section. Follow-up Mann-Whitney tests indicated that all pairwise differences were significant with p<0.001, except the difference between the two recovery process types involving explanation (first explanation step immediately after detection versus at a later stage), p=0.220. Logically, the processes involving only detection had fewer steps than all other process types. Processes with detection followed by only countermeasures had fewer steps than both process types that also involved explanation.

Additional details

Now that the accuracy of the recovery process phase model with regard to the possible steps and transitions has been confirmed and a closer look has been taken at the different forms or types of recovery processes encountered in the collected data, only part *c)* of research question 1 remains to be answered. This part concerned further specification of the goals of each of the recovery process phases and therefore required a more in-depth look at the goals of the steps involved in recovery processes about which I had collected data. While there were no problems, as was stated before, in assigning the steps involved in the recovery processes about which data was collected to the categories of either detection, or explanation, or countermeasures, it became clear that especially within these last two categories, different sub-goals could be distinguished. Johannsen (1988) already suggested that people may strive for different goals in the correction phase. In this regard he offered the example of compensation for failures, when the preliminary goal is to continue with degraded performance, in systems where operations have to be continued simultaneously. Mo & Crouzet (1996) and also Kontogiannis (1999) distinguished three possible corrective goals in recovery processes: backward, forward, and compensatory recovery. In backward recovery the system is brought back to its state from prior to the occurrence of the failure. In forward recovery the system is brought to an intermediate state via which time is bought to find and implement a better solution. In compensatory recovery redundant equipment is activated to bring the system to the originally desired state.

While the distinctions made by these researchers were definitely useful, an in-depth investigation of the goals associated with the recovery steps involved in the collected cases revealed that these distinctions did not fully cover all the different goals encountered in the dataset. My findings suggested two possible sub-goals for the explanation phase and four for the countermeasures phase. The resulting seven phases and sub-phases are described below, illustrated with examples from the three initial studies:

- Detection of the problem situation (D). In this phase the person or persons involved become aware of the presence of the problem, i.e. the deviation. At this moment all the relevant aspects and the causes of the deviation are not necessarily known to the person(s) involved. An example of this phase is a chemical process plant operator noticing a high pressure alarm in the control panel.
- Explanation – definition of the problem (EDP). In this sub-phase of the explanation phase the person(s) involved gather all the relevant information about the deviation or problem situation, to specify what exactly is going on, or how the deviation manifests itself, without looking for the causes at this moment. This sub-phase usually occurs early on in the sequence of recovery steps, just after detection of the deviation. In cases where urgent countermeasures are needed first, the problem definition phase may be postponed a while, if the person(s) involved feel that relatively much time is needed to arrive at a proper and workable problem definition. It may also (re-) occur at a later moment, for example if countermeasures that have been taken have changed the situation and a new assessment is needed of what the remaining deviation is. Or if newly found insight into the causes of the deviation sheds a new light on the deviation's extent. An example of the EDP phase is when the chemical process plant operator (after having noticed the high pressure alarm in the detection phase) checks the pressures in other parts of the installation as well to find out the extent of the deviation, or checks the temperature in the affected section to gain more information.

- Explanation – identification of the causes (EIC). In this sub-phase of the explanation phase, the person(s) involved in the recovery process try to find out what caused the deviation of problem situation. Knowing the deviation's cause or causes may be necessary before the appropriate countermeasures can be identified. In other cases there is no such need, but even then a search for the causes may still be conducted at some point, as part of an incident investigation or other organizational learning process. For familiar deviations, or situations where the connection between deviation and cause is very obvious, the causes (or failures involved) may be known almost immediately and no explicit search for them is required. In other cases the causes are more obscure or well hidden, or the deviation is the result of a complicated mix of failures and other contributing factors. In those situations finding the causes takes more time and effort. An example of the EIC phase is when the chemical process plant operator inspects the pump involved in the high pressure alarm situation for potential causes, such as valves in closed positions which should be open, blockages, and so on.
- Countermeasures – stabilization (CS). Stabilization of the situation is a sub-phase of the countermeasures phase. CS may be necessary in order to prevent things from getting further out of hand; to stop the process of otherwise continually increasing negative effects, such as damage, spills, or process deviations. These steps are not aimed at reducing the negative effects that have already occurred, they simply 'freeze' (not counter) the problem. This CS phase is applicable for special types of situations, where it is either not possible at all to initiate other counteractions without taking those stabilizing steps first, or where other counteractions will not be as effective as they could be until the situation is stabilized (like mopping while the tap is still running). One of the aims of stabilization can simply be to gain time for additional steps. A stabilized situation is not the same as the normal situation, additional actions or processes are necessary to reach that state. An example of the CS phase is when a chemical process plant operator stops a leaking pump so that the product spill resulting from the leakage does not increase any further.
- Countermeasures – mitigation (CM). Another countermeasures-sub-phase is mitigation, aimed at actively reducing the amount, scope and/or impact of the negative consequences to which the preceding failure(s) have already led at the moment at which the mitigation steps are started. This is more than simply stopping the further deterioration of the damage process (i.e. stabilization); it also involves damage control and reduction. This phase may be applicable for situations where a return to the normal situation from before the occurrence of the problem is either not possible at all, or at least not within a foreseeable time; where it is obvious that there will be some remaining damage or other negative consequences after the mitigation steps have been taken. The aim of this type of countermeasures is to at least limit these consequences as far as possible. Additional steps are still needed after mitigation for a complete

return to the normal situation, that is, if such a return is possible at all. An example of mitigation is extinguishing a fire. The result after the flames are put out is still a situation in need of additional repairs or replacements.

- Countermeasures – temporary correction of the problem (CTCP). Another sub-phase of the countermeasures phase involves temporary corrections of the problem at hand. Temporary corrections are not intended to last forever. They involve a 'quick fix', aimed at returning to a normal, workable situation at least for a while, with steps that can be completed in a relatively short time. Temporary repairs (e.g. using tape to patch up a leak in a pipeline) and use of spare capacity (such as spare pumps) are examples of steps belonging to the CTCP phase. Generally, after a temporary correction, additional, more permanent countermeasures are still needed, either because the temporary correction is not durable enough to keep the problem solved forever, or because it uses normally spare capacity which should be returned to that status.
- Countermeasures – permanent correction of the problem (CPCP). A final sub-phase of the countermeasures phase involves permanent or lasting corrections of the problem at stake, which are aimed at returning to a normal situation, without remaining consequences in the installation, process or personnel. The corrective measures implemented as part of the CPCP phase have a lasting effect; it doesn't wear off over time like the effect of temporary corrections would. An example of the CPCP phase is when the chemical process plant operator cleans and reinstalls the filter in the pump that was causing the high pressure because of dirt caught in it.

A graphical representation, similar to the higher-level phase model presented in figures 4 and 5, but containing the newly distinguished sub-phases as well, would result in a crisscross of lines representing all the possible transitions from each of the explanation- and countermeasures sub-phases to each other, including themselves. For this reason, and assuming that one can imagine what this jumble would look like, I have not included such a figure here.

As may be obvious from the definitions and further explanations given above, with the exception of the detection phase, not all of these (sub-)phases are equally likely to occur in any given recovery process. Table 6 shows how often each of these seven types of recovery steps occurred within the events on which data was collected.

Table 6 Frequencies of occurrence of each of the more detailed recovery process phases, for each of the studies (and number of cases where phase occurred at least once)

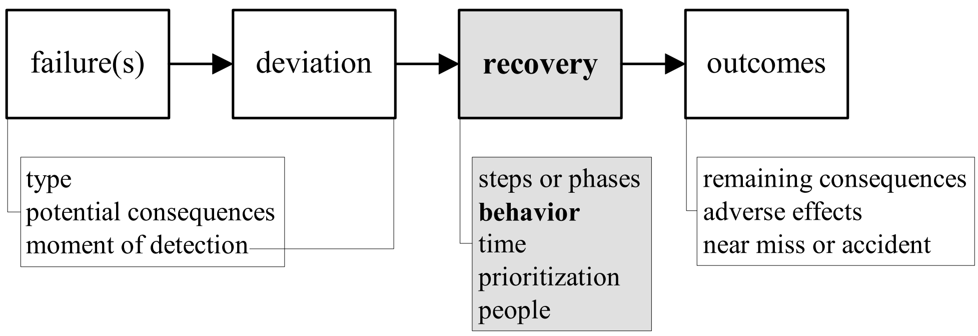
recovery process (sub-)phase	**number of occurrences of phase (number of cases in which phase occurred at least once)**					
	study 1		**study 2**		**study 3**	
D	50	(50)	33	(33)	39	(39)
EDP	20	(19)	9	(9)	11	(10)
EIC	38	(29)	14	(14)	22	(17)
CS	6	(6)	5	(5)	13	(13)
CM	5	(5)	2	(2)	14	(13)
CTCP	15	(13)	0	(0)	10	(9)
CPCP	104	(47)	51	(31)	65	(37)
total	*238*		*114*		*174*	

Table 6 shows that from the two different types of explanation sub-phases, EIC, where the causes of the deviation are identified, occurred more often than the EDP phase, where more information about the deviation is collected without actually looking for its causes. In the cases where an EIC phase occurred and no EDP steps were taken, the extent of the deviation was already obvious (or otherwise known) at the moment of detection. From the different types of countermeasures sub-phases, CPCP was by far the most frequent, i.e. the permanent corrections. Regularly, more than one CPCP step was needed in the recovery events about which data was collected. Even though there were not that many examples of CS and CM, i.e. stabilization and mitigation steps, those cases where these steps did occur showed that CS and CM phases usually occur in early stages of the recovery process, to make further steps possible. In study 2, involving only cases where self-made errors were recovered, CTCP steps, i.e. temporary corrections, were never needed, indicating that in these situations people never had to wait long before more permanent corrections could be implemented.

Recall from tables 4 and 5 that the cases in studies 1 and 3 on average involved more steps than those in study 2. To find out if this difference in number of steps could be attributed to any of the sub-phases following after detection in particular, I analyzed the adjusted standardized residuals for the cells of table 6 for all of the E and C sub-phases. I left the detection phase outside the analysis, since this phase by default occurred exactly once in every case in each of the three studies. The analysis indicated that the only sub-phase in study 2 that occurred significantly less frequently than expected based on the table's marginal totals was CTCP (this phase actually never took place), but on the other hand, CPCP occurred significantly more often than expected in that study. Based on these findings it seemed fair to say that the smaller amount of steps in cases in this study was not concentrated in any particular phase except possibly the CTCP phase. In addition, the analysis revealed that study 1 involved significantly fewer CS and fewer CM steps than expected, but there was no clear indication that any particular phase was responsible for the larger number of steps found in this study. The analysis also revealed that study 3 involved significantly more CS and CM steps than expected, so these may have contributed to the larger number of steps. On the other hand, study 3 involved significantly fewer CPCP steps than expected. In the analysis where all countermeasures sub-phases were combined into one category of countermeasures and all explanation sub-phases were combined into one category explanation, no significant adjusted standardized residuals were found.

To avoid difficulties in statistical tests resulting from too few observations per variable level, such as large numbers of empty cells in cross-tabulations, for most of the analyses involving recovery process phases presented in the rest of this chapter only the high-level phases detection, explanation and countermeasures were used. For some of the research questions, however, specific expectations existed with regard to certain sub-phases. Where this was the case, the corresponding findings are also presented.

3.3.2 Behavior involved in recovery steps



The second research question concerned the types of behavior and cognitive processes involved in recovery: *For each of the distinguished types of recovery process phases (i.e. detection, explanation, and countermeasures), a) which proportion of all the steps belonging to that phase involves ad-hoc behavior and which proportion pre-planned behavior, and, b) furthermore, which proportion involves skill-based behavior, which proportion rule-based behavior, and which proportion knowledge-based behavior?*

Ad-hoc versus pre-planned

With regard to part *a)* of research question 2, as was explained in subsection 2.2.2, the differentiation between pre-planned and ad-hoc (i.e. unplanned) behavior is made from the organization's point of view. More pre-planned behavior indicates a standardized preparation for problem situations, through formal or informal procedures or through training, which can be helpful, but not necessarily for all situations. For ad-hoc steps, organizations still need to rely on their employees' creative thinking and problem solving abilities, requiring a different kind of support than pre-planned provisions. My expectations with regard to the types of behavior involved in recovery in high reliability organizations were as follows: The detection of failures is often facilitated by pre-planned checks, warning systems, alarms, or other defenses or barriers, especially for known failures, which should be the majority of all the possible failures. The countermeasure phases are expected to involve predominantly pre-planned steps, which may especially be true for the more urgent steps aimed at mitigation and stabilization, and otherwise a combination of planned and ad-hoc steps. Explanation of the deviation and its causes is very situation-specific and hard to pre-plan, so even if tools or procedures exist to support the diagnosis, they still need to be tailored to the situation. Even though high reliability organizations may have provided such support, I expected predominantly improvised, ad-hoc steps during explanation.

For each of the three studies, table 7 contains an overview per recovery process phase of the numbers of planned (i.e. pre-planned) and unplanned (i.e. ad-hoc) steps and steps where both types were combined. Also their relative contributions are indicated, i.e. with respect to all the steps belonging to that phase.

Table 7 Numbers of planned, unplanned, and combined type of steps, per recovery process phase (and percentages of all steps belonging to that phase), for each of the studies

		number of steps (percentage of all steps belonging to that phase)		
		study 1	**study 2**	**study 3**
detection	unplanned	21 (42%)	17 (51.5%)	26 (66.7%)
	both types	18 (36%)	7 (21.2%)	6 (15.4%)
	planned	11 (22%)	9 (27.3%)	7 (17.9%)
	not certain	0 (0%)	0 (0%)	0 (0%)
explanation	unplanned	44 (75.9%)	11 (47.8%)	22 (66.7%)
	both types	10 (17.2%)	6 (26.1%)	8 (24.2%)
	planned	3 (5.2%)	6 (26.1%)	3 (9.1%)
	not certain	1 (1.7%)	0 (0%)	0 (0%)
countermeasures	unplanned	29 (22.3%)	12 (20.7%)	32 (31.4%)
	both types	35 (26.9%)	6 (10.3%)	50 (49.0%)
	planned	59 (45.4%)	40 (69.0%)	20 (19.6%)
	not certain	7 (5.4%)	0 (0%)	0 (0%)

While I expected a lot of pre-planned steps or combinations of pre-planned and ad-hoc steps in the detection phase, table 7 shows that in all studies actually quite a large amount of unplanned, ad-hoc steps were found for this phase. In ad-hoc detection, the people involved detected things that were wrong even though there were no formal or informal protocols, procedures, or instructions that suggested they should be looking for any such thing. An example was the detection done by a field operator who heard that a pump he walked past made an abnormal sound. When I asked this operator if all other operators in the same job position would have been able to notice this, he told me that he did not think so, this was not something they were trained in, and no procedures told the operators to do what he did, and it had taken him over twenty years of experience to be able to detect such a minor change in sound as he had done.

The majority of countermeasures indeed involved pre-planned steps, with occasionally, especially in study 3, a mixture of both types of steps, and sometimes also unplanned, ad-hoc steps. Examples of pre-planned countermeasures were all actions that were exactly described in procedures for dealing with specific deviations, such as exchanging a blocked filter. Examples of the combined type of countermeasures, partly pre-planned and partly ad-hoc, were all actions that were to some extent supported by provisions made by the organization, such as actions based on procedures that were first tailored to cope with some novel aspects of an otherwise common deviation, or actions that were devised in an ad-hoc manner, but whereby already available tools were used, or the process control computers. An example of an ad-hoc countermeasure was the action performed by an operator who detected an unprotected piece of scaffolding pipe sticking out in a cage ladder: since it would take some time to process the work order (pre-planned countermeasure) he wrote to get the situation properly corrected, in the mean time, he wrapped some padding material around the pipe end, and marked and fixated it with fluorescent tape, so people would easily notice the hazard and would not get hurt when they would accidentally hit it when climbing the ladder. Ad-hoc or unplanned countermeasures are not necessarily performed without a plan: even though no plan is provided by formal or informal procedures or practices, an action plan may still be devised by the persons involved to guide the implementation of their countermeasures.

As predicted, the steps aimed at explanation of the deviation and its causes involved mostly unplanned, or at least partly unplanned steps. Explanation phases that were partly planned and partly unplanned were found, for example, in the cases where instrumentation engineers used diagnostic equipment provided by the organization to locate the causes of a failing indicator – a checking process for which no detailed procedures are prescribed as the necessary steps are unique for the specific situation involved.

In total, 11 of the steps aimed at stabilization, the CS sub-phase, were pre-planned, four involved a combination of both types, and only three were ad-hoc. With regard to the steps aimed at mitigation, the CM sub-phase, in total, eight steps were pre-planned, another eight involved a combination of both types, and only three were ad-hoc. These findings confirmed the expected predominance of pre-planned steps for the CS and CM sub-phases.

Skill-, rule-, or knowledge-based

With respect to part *b)* of research question 2, the distribution of the steps per recovery process phase across the categories skill-based, rule-based, and knowledge-based behavior is shown in table 8.

Recall from subsection 3.2.5, where the variables and data coding process were described, that for steps involving several levels of information processing behavior, for example both knowledge-based and rule-based behavior, the entire step was coded as involving the highest of the applicable levels of behavior.

Table 8 Numbers of skill-, rule-, or knowledge-based steps, per recovery process phase (and percentages of all steps belonging to that phase), for each of the studies

		number of steps (percentage of all steps belonging to that phase)					
		study 1		study 2		study 3	
detection	skill-based	20	(40%)	15	(45.5%)	25	(64.1%)
	rule-based	9	(18%)	5	(15.2%)	3	(7.7%)
	knowledge-based	21	(42%)	13	(39.4%)	11	(28.2%)
	not certain	0	(0%)	0	(0%)	0	(0%)
explanation	skill-based	0	(0%)	0	(0%)	3	(9.1%)
	rule-based	5	(8.6%)	9	(39.1%)	2	(6.1%)
	knowledge-based	52	(89.7%)	14	(60.9%)	28	(84.8%)
	not certain	1	(1.7%)	0	(0%)	0	(0%)
countermeasures	skill-based	4	(3.1%)	19	(32.8%)	23	(22.5%)
	rule-based	67	(51.5%)	27	(46.6%)	16	(15.7%)
	knowledge-based	51	(39.2%)	12	(20.7%)	63	(61.8%)
	not certain	8	(6.2%)	0	(0%)	0	(0%)

The observations with regard to the unexpected, important role of unplanned, ad-hoc detection were confirmed when looking at the human information processing behavior level involved in the detection phase: respectively 42%, 39.4% and 28.2% of the detection phases involved knowledge-based behavior. An example of knowledge-based detection was the case where an operator who, while reading a procedure in preparation of a not very familiar task, found out that the procedure referred to a timer that had actually never been installed. To detect this, the operator used his knowledge of the plant, verified this with co-workers, and even checked the actual situation in the plant once more to be certain. Ad-hoc detection regularly involved skill-based behavior as well, such as in the example given earlier of the operator detecting an abnormal sound. Obviously, pre-planned detection can be skill-based too, for example in cases where detection is triggered by an auditory or visual alarm. An example of rule-based detection was the detection made by an operator who needed a compressed air cylinder for a task for which the use of a breathing apparatus was prescribed: following standard procedures, he first checked the pressure level of the sealed, so supposedly full compressed air cylinder he had picked up from the on-site storage, and found out that this was only half of the normal level, so that the supply would only last half as long as would be expected.

The explanation phases involved almost always knowledge-based behavior, occasionally rule-based, and rarely skill-based. An example of knowledge-based explanation was the case where a lab technician found unexpected test results for a blend he had just prepared: he was able to establish what caused these results by tracing back the steps he had taken during the preparation. By doing this, he discovered that he had used different proportions of the materials combined in the blend than what he originally thought. Rule-based explanation occurred, for example, in cases where people followed checklists or flow diagrams specifically developed to establish the cause of certain deviations. Skill-based explanation only occurred when further information about the extent or the causes of the deviation were so obvious that they could not be missed even without looking for them. An example was the case where a drain was left open in a line-up that had just been made. When the operators started the product flow through this line-up they heard a hissing sound coming from the drain. The connection of this sound to the oversight made in the line-up was easily made, without any further thinking or following rules.

The countermeasures phases were dominated by rule-based behavior, confirming the importance of pre-planned steps, while knowledge-based behavior also played an important role, especially in study 3. Examples of rule-based countermeasures were the recovery steps taken by an operator to replace a defective lock-ring, after he had identified this ring as the cause of a leakage in the circulation pipeline of a pump. Sometimes, countermeasures could even be performed at skill-based level, if they concerned routine behavior for the person involved. An example of a skill-based countermeasure was the stabilization action performed by a control room operator who had made a typing error in entering the temperature setpoint for a furnace, which he detected immediately after entering the setpoint: at that moment, without further thinking, he immediately switched to manual control, to avoid setting off all kinds of automatic plant responses. As soon as countermeasures had to be implemented for deviations with a few novel characteristics that were not covered by existing procedures or routines, or for completely unfamiliar deviations, behavior at knowledge-based level was necessary.

In total, 14 of the steps aimed at stabilization, the CS phase, involved skill-based behavior, two involved rule-based behavior, and eight knowledge-based behavior. With regard to the mitigation steps, the CM phase, seven of those involved skill-based behavior, 6 rule-based, and seven knowledge-based. So, while roughly half of these CS and CM steps involved almost automatical, often well-rehearsed responses to the situation, obviously there was also an important role for higher levels of information processing behavior.

Link

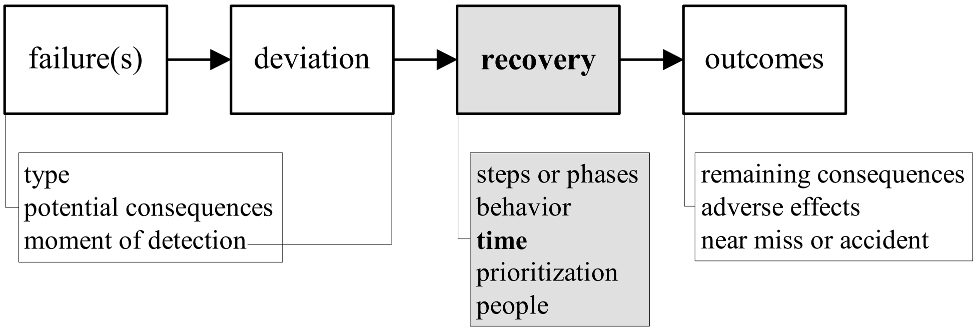
As can be seen from table 9 below, a link existed between the level of human information processing behavior involved and the planned or unplanned character of the steps: pre-planned steps were almost always rule-based (procedure or protocol is followed), but could be skill-based too if the step was very familiar, and unplanned, ad-hoc steps were mostly either knowledge-based or skill-based, and only occasionally rule-based, depending on the amount of reasoning involved.

Table 9 Number of steps as a function of the extent to which they were pre-planned and the information processing behavior level involved, per recovery process phase, for each of the studies

		number of steps		
phase	**'pre-plannedness'**	**skill-based**	**rule-based**	**knowledge-based**
detection	unplanned	35	3	26
	both types	8	4	19
	planned	17	10	0
explanation	unplanned	2	1	74
	both types	1	5	18
	planned	0	10	2
countermeasures	unplanned	6	2	49
	both types	10	14	42
	planned	23	64	4
all phases combined	unplanned	43	6	149
	both types	19	23	79
	planned	40	84	6

Table 9 also shows that the predominant information processing behavior type in ad-hoc detection was skill-based, but there was also an important role for knowledge-based behavior. For unplanned countermeasures, but even more so for the unplanned steps in the explanation phase, the roles of skill-based and knowledge-based behavior were reversed, knowledge-based behavior was far more frequent here than skill-based.

3.3.3 Time spent on recovery steps and process



The third research question concerned the time involved in recovery steps and the complete recovery process: *How long do a) different types of recovery steps at various stages in the process, and b) the entire process take?*

Separate recovery steps

For each of the three studies, table 10 shows how all of the separate recovery steps were distributed across the different categories of amounts of time spent on the step.

Table 10 Number of recovery steps in each recovery process phase per category of amount of time spent on these steps, for each of the studies

time spent on separate recovery steps	**number of recovery steps**								
	study 1			**study 2**			**study 3**		
	D	**E**	**C**	**D**	**E**	**C**	**D**	**E**	**C**
up to 5 minutes	49	28	46	29	18	49	39	15	39
between 5 min. and 1 hour	1	28	66	4	4	7	0	16	37
1 to 8 hours	0	0	16	0	0	2	0	1	21
8 to 16 hours	0	0	0	0	0	0	0	1	1
16 to 24 hours	0	0	0	0	0	0	0	0	0
1 to 7 days	0	0	0	0	1	0	0	0	2
1 to 4 weeks	0	0	0	0	0	0	0	0	1
even longer than 4 weeks	0	0	0	0	0	0	0	0	1
not certain	0	2	12	0	0	0	0	0	0
total	*50*	*58*	*130*	*33*	*23*	*58*	*39*	*33*	*102*

Concerning the amount of time spent on separate recovery steps, both with regard to the steps belonging to the countermeasures phase and explanation steps differences were found between the studies: on average, these tended to take longer in the cases from studies 1 and

3 than from study 2. In all three studies, the detection phase hardly ever took more than five minutes (only in five out of 122 cases). Of all the recovery steps involved in the three studies, only 44 steps belonging to the countermeasures phase and one explanation step took longer than one hour to complete. Generally, less time was spent on the steps taken at early stages during the recovery processes in the dataset than on steps taken at later stages. This was not only so for detection, logically always the first step in recovery, but also for early steps aimed at explanation, temporary corrections, stabilization, and mitigation. Especially stabilization and mitigation were typically performed quickly early on in the process, to buy time for additional steps.

Kruskal-Wallis tests to evaluate differences in amounts of time spent on a step between ad-hoc or unplanned, pre-planned, and steps involving a combination of both types of behavior showed insignificant results for detection steps, χ^2 (2, N=122)=1.634, p=0.604, and for steps aimed at explanation, χ^2 (2, N=111)=0.805, p=0.692. Only with regard to steps belonging to the countermeasures phase, were significant differences found in the time spent on the step between steps involving ad-hoc, pre-planned, and combinations of both types of behavior, χ^2 (2, N=278)=22.786, p<0.001. Follow-up Mann-Whitney tests for this phase indicated that both the difference between ad-hoc and pre-planned steps and between the combined type and pre-planned steps were significant, respectively z=-2.736, p=0.006 and z=-4.682, p<0.001. The pre-planned countermeasures took less time than the ad-hoc countermeasures and also less time than the countermeasures where both pre-planned and ad-hoc actions were combined.

I also conducted Kruskal-Wallis tests to evaluate differences in amounts of time spent on a step between skill-based, rule-based, and knowledge-based steps. For detection steps, the test result was significant, χ^2 (2, N=122)=8.848, p=0.023. Follow-up Mann-Whitney tests to evaluate pairwise differences among the three behavior types indicated that only the difference between time spent on skill-based detection and time spent on knowledge-based detection was significant, z=-2.633, p=0.012, with less time spent on skill-based detection. Also for steps aimed at explanation, the test result was significant, χ^2 (2, N=112)=7.326, p=0.019. Follow-up Mann-Whitney tests indicated that only the difference between time spent on rule-based explanation and time spent on knowledge-based explanation was significant, z=-2.184, p=0.033, with less time spent on rule-based explanation. For steps belonging to the countermeasures phase, the test result was significant as well, χ^2 (2, N=279)=57.578, p<0.001. Follow-up Mann-Whitney tests indicated that all of the three pairwise differences were significant: z=-5.358, p<0.001 for skill- versus rule-based countermeasures, z=-7.180, p<0.001 for skill- versus knowledge-based countermeasures, and z=-3.684, p<0.001 for rule- versus knowledge-based countermeasures. In summary, the least time was spent on skill-based countermeasures and the most on knowledge-based countermeasures, with time spent on rule-based countermeasures averaging somewhere in between.

Entire recovery process

The total amount of time passed before a recovery process is completed is not only the sum of the time spent on each of the steps involved, it also includes the periods between the recovery steps, when no steps are taken towards further recovery. Those periods between recovery steps may be the result of circumstances whereby other, unrelated tasks are more urgent and important, or it may not yet be possible to implement the required recovery step. This explains why a recovery process can still take a long time to complete, even when the individual steps involved each are completed in a relatively short time. Table 11 shows how the recovery processes collected in the three studies were distributed across the different categories of total amounts of time passed between process start and completion.

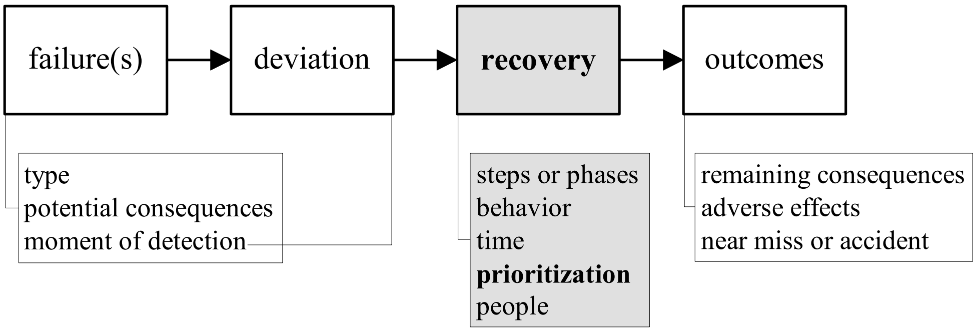
Table 11 Number (and percentage) of cases per category of total amount of time passed between recovery process start and completion, for each of the studies

time passed from recovery process start until completion	**number (percentage) of cases**					
	study 1		**study 2**		**study 3**	
up to 5 minutes	1	(2%)	16	(48.5%)	8	(20.5%)
between 5 minutes and 1 hour	9	(18%)	11	(33.3%)	7	(17.9%)
1 to 8 hours	11	(22%)	3	(9.1%)	3	(7.7%)
8 to 16 hours	0	(0%)	0	(0%)	0	(0%)
16 to 24 hours	4	(8%)	1	(3.0%)	2	(5.1%)
1 to 7 days	11	(22%)	1	(3.0%)	8	(20.5%)
1 to 4 weeks	4	(8%)	0	(0%)	6	(15.4%)
even longer than 4 weeks	10	(20%)	1	(3.0%)	5	(12.8%)
total	*50*	*(100%)*	*33*	*(100%)*	*39*	*(100%)*

A Kruskal-Wallis test revealed significant differences between the three studies with regard to the total amount of time passed between the recovery process start and completion, χ^2 (2, N=122)=31.925, p<0.001. Follow-up Mann-Whitney tests indicated that both the difference between studies 1 and 2 and the difference between studies 2 and 3 with regard to time passed from process start to completion were significant, respectively z=-5.786, p<0.001 and z=-3.796, p<0.001. Recovery processes in cases involving self-made errors took on average significantly less time before they were completed, than in cases involving other types of failures. Another Kruskal-Wallis test revealed that there were no significant differences with regard to the total amount of time passed from recovery process start until completion between the four recovery process types that were distinguished in subsection

3.3.1, χ^2 (3, N=122)=1.989, p=0.603. A Mann-Whitney test to identify differences with regard to this total amount of time between cases with and cases without explanation was not significant either, z=-0.999, p=0.327. The Spearman correlation between the total amount of time passed from recovery process start until completion and the total number of steps involved was $r_s(120)$=0.427, p<0.001. This confirmed what logically might be expected, that processes with more steps took a longer time to complete than processes involving less steps.

3.3.4 Prioritization



The fourth research question involved prioritization: *Based on which factors do people prioritize tasks when a deviation needing recovery presents itself at work?* In subsection 2.2.4 the presence of other tasks requiring immediate attention, the urgency and importance of recovery steps, and the necessary time and means were proposed as important factors in prioritizing tasks when people are confronted with a situation in need of recovery.

For each of the three studies, table 12 shows how the cases were distributed over the different categories of extents to which there were other tasks requiring immediate attention.

Table 12 Number of cases per category of amount other tasks requiring attention when recovery is needed, for each of the studies

	number of cases		
amount of other tasks requiring attention	**study 1**	**study 2**	**study 3**
all other tasks can wait	31	30	27
some other tasks also require immediate attention	18	3	11
many other tasks also require immediate attention	1	0	1
total	*50*	*33*	*39*

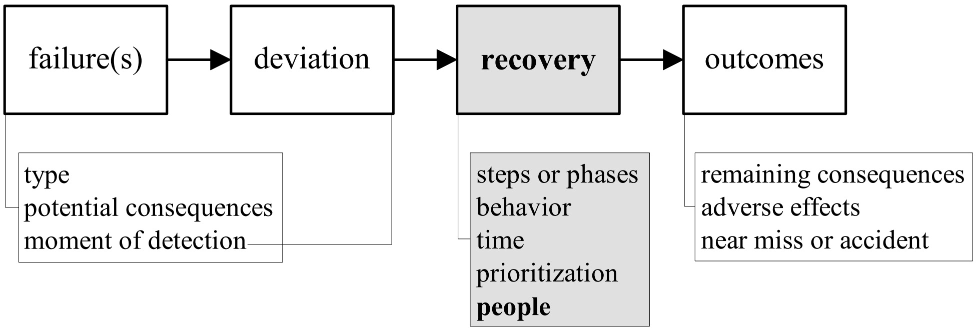
A Kruskal-Wallis test revealed significant differences between the three studies with regard to this presence of other tasks requiring attention, χ^2 (2, N=122)=8.468, p=0.014. The follow-up Mann-Whitney tests showed that the differences between studies 1 and 2 and studies 2 and 3 were significant, respectively z=-2.915, p=0.004 and z=-2.261, p=0.036: in the cases in study 2 there were significantly less situations where other tasks also required immediate attention when recovery steps were needed than in the cases in the other two studies. In the cases where there were other tasks requiring immediate attention, people made sure that the problem would at least be more or less contained, before continuing with other tasks, often by taking a few initial steps without already completing the entire process.

Recall that the urgency of recovery steps corresponds with the time available until the earliest possible occurrence of potential consequences if no recovery steps would be taken, and the importance of recovery steps with the type and severity of the potential consequences. Since these two factors are inherent to the failures involved in the situations in need of recovery and in 3.3.6 an entire subsection is devoted to this topic, I will hold off on a detailed discussion of these factors until that subsection. For the moment, here I will only present results relevant for the question how people prioritize their tasks.

At the time of the data collection in the initial three studies, the factors ‘necessary time’ and ‘means’ were mainly viewed as reasons to involve others (see subsection 3.3.5). As a result, from these studies I only have anecdotal evidence of the role these factors played in prioritization. I did not systematically record and code information regarding, for example, possible delays in the recovery process as a consequence of not having the means or time needed for recovery.

Unfortunately, the information I had recorded regarding the time passed between detection and the first recovery step after detection was not detailed enough to allow structured coding and analysis. The only measure available in the dataset that could be used as an indication of the priority given to recovery was the total time passed from the start of the recovery process until its completion (see subsection 3.3.3, table 11). To get an indication of the effects on this total amount of time of the presence of other tasks requiring urgent attention, of the time available until the first possible occurrence of potential consequences, and of the type and severity of potential consequences (combined in a summary score, see subsection 3.3.6), Spearman correlations were calculated. These correlations were respectively $r_s(120)$=0.532, p<0.001, $r_s(72)$=0.057, p=0.628 (i.e. not significant), and $r_s(120)$=0.239, p=0.008. This indicated that when more other tasks also required attention, and for more and more severe potential consequences, recovery processes tended to take longer. The Spearman correlation between the extent to which other tasks were present that also required attention and the total number of steps in the recovery process, $r_s(120)=0.264$, $p=0.004$, indicated that more separate recovery steps were taken during recovery processes when tasks other than recovery also needed attention than during processes when this was not the case.

3.3.5 People involved in recovery



The fifth research question concerned the people who play a role in recovery processes: *If others than those already involved become involved in a recovery process at work: a) what are the organizational roles of these people, b) when do they become involved, c) for what reasons do they become involved, and d) what is the effect of their involvement? Furthermore, if those involved in detection do not involve any others in the rest of the recovery process at all, for what reasons does this happen?*

Number of people

It may be useful to establish first whether there actually were processes in which several people played a role in each of the three studies. From table 13, in which the distribution of the recovery processes over the different total numbers of people involved is presented, it is obvious that in most processes more than just one person was involved.

Table 13 Number of cases per total number of people involved in the recovery process, for each of the studies

total number of people involved in recovery process	number of cases		
	study 1	study 2	study 3
1	2	15	8
2	5	12	4
3	8	3	4
4	4	0	4
5	6	2	3
6	6	1	6
7	6	0	1
8	11	0	2
9	0	0	1
10 or more	2	0	6
total	*50*	*33*	*39*

A Kruskal-Wallis test revealed a significant difference between the three studies with regard to the total number of people involved, $\chi^2(2, N=122)=35.975$, $p<0.001$. The follow-up Mann-Whitney tests showed that the differences between studies 1 and 2 and between studies 2 and 3 were significant, respectively $z=-6.131$, $p<0.001$ and $z=-4.078$, $p<0.001$: less people are involved in cases where self-made errors are recovered than in other cases. The correlations between total number of recovery process steps and total number of people involved (Pearson's correlation) and between total amount of time passed from recovery process start to completion and total number of people involved (Spearman correlation) were $r(120)=0.392$, $p<0.001$, and $r_s(120)=0.611$, $p<0.001$, respectively. This confirmed what logically might be expected, that in longer processes, both with respect to number of steps and total amount of time, more people tended to be involved. The Spearman correlation between the extent to which other tasks were present that also required attention and the total number of people in the recovery process, $r_s(120)=0.417$, $p<0.001$, indicated that more people were involved in recovery processes where tasks other than recovery also needed attention than in processes where this was not the case.

Organizational roles of those involved

The organizational roles of the persons involved in recovery (part *a)* of research question 5) varied from people within the operators' own team including their supervisor, to electrical equipment & instrumentation engineers, maintenance engineers, people in management positions within the organization, on-site, permanently hired contractors specializing in certain tasks, and occasionally specialists or suppliers from outside the organization. Most of the recovery processes in studies 1 and 3 could be solved within the organization. Only in a few cases a specialist (e.g. to deal with instrumentation or automation) or supplier from outside the organization (e.g. for new parts replacing broken ones) was involved in the recovery process, and only in permanent corrective actions. Quite regularly, people from different departments and also on-site contractors played a role in the recovery processes. In study 2, on the other hand, assistance from outside the organization was never needed, and only rarely people from a different department had to be involved to complete the recovery process.

When did others become involved

With regard the process stages *when* these different people were involved, part *b)* of research question 5, a distinction can be made between people involved in detection, and people involved in the rest of the process. In study 1, detection was always done by people other than those involved in the failures from which recovery was needed. The errors in study 2 were detected by the exactly the same persons as those who committed them in 19 out of 33 cases (57.6%). In 5 cases (15.2%) a different number of people than those who made the error were involved in its detection, but still at least one person was involved in both the error and detection. In 9 cases (27.3%), the error was detected by a different person or different persons than the error producer(s). Still, in all of those cases the error producer(s) contributed to its recovery, in a later phase of the recovery process, after having been informed about the error by the person who originally detected the error. In study 3, in 29 out of 39 cases (74.4%), people other than those involved in the failures from which recovery was needed were the ones who detected that something was wrong. In 6 cases (15.4%), the same persons as those involved in the failures also detected them, and in another 4 cases (10.3%), even though a different number of people were involved in the failures than in the detection, at least one person was involved in both.

Regardless of whether those involved in the preceding failures were the same or different people as those involved in the recovery process, even during the recovery process different people may become involved. The first moment when additional people (i.e. more than just one) became involved in the recovery process differed significantly between the three studies. There was a difference between the studies with regard to the type of step or phase involved, $\chi^2(4, N=97)=10.305$, $p=0.029$ (Monte Carlo estimation for exact significance): in the detection phase in 6, 4, and 7 cases, respectively in study 1, 2, and 3, in an explanation phase respectively in 15, 11, and 8 cases, and in a countermeasures phase respectively in 27, 3, and 16 cases. Analysis of the adjusted standardized residuals indicated that in study 1, the phase when for the first time additional people became involved was significantly more often than expected a countermeasures phase. In study 2, this was significantly more often than expected an explanation phase and significantly less often than expected a countermeasures phase. In study 3, the distribution across the phases with regard to the first phase when more persons become involved was not significantly different than expected. A Kruskal-Wallis test indicated that there was also a difference between the studies with regard to the step number in the sequence of all consecutive recovery process steps in which for the first time more people became involved, $\chi^2(2, N=97)=7.387$, $p=0.025$. Follow-up Mann-Whitney tests showed that only the difference between studies 1 and 2 was significant, $z=-2.595$, $p=0.010$: in study 2 others became involved earlier in the sequence of recovery steps than in study 1. After the first moment additional people became involved, many recovery processes also involved additional moments at later stages in the process when again other people became involved, especially the processes consisting of larger numbers of steps.

Why did others become involved

With regard to the reasons *why* these different people became involved in the recovery process when they did, part *c)* of research question 5: these reasons were recorded for every moment in the recovery processes when the assistance of others was sought or a complete handover occurred. As was pointed out already in subsection 2.2.5, multiple reasons could apply to any such moment. In the cases where the first moment when others became involved was immediately after detection, 41 cases in total, the predominant reasons given for this involvement were either what procedures prescribed with regard to who is allowed to perform the required task, or that the person(s) involved in detection lacked the required knowledge or expertise to continue the recovery process by themselves, or regularly even both were applicable. In table 14 an overview is presented of the frequencies of the different reasons given for all of the assistance recruitments and handovers that occurred throughout all of the collected recovery processes.

Table 14 Number of times specific reasons for seeking assistance or handing over recovery tasks applied, for each of the studies

reason	number of times that reason was mentioned		
	study 1	study 2	study 3
person is not allowed by procedures to perform required task	65	10	36
person does not have required or sufficient knowledge / expertise	25	10	32
person does not have necessary tools / equipment	16	3	20
person does not have sufficient human resources alone	5	6	10
person does not have time now	2	0	0
step not possible yet but only at a later moment	1	0	1
coincidence	8	1	7
others need to be informed so potential additional problems are prevented	4	3	8
other than any of the above	3	3	3

Table 14 shows that the two predominant reasons for involving others immediately after detection continued to play an important role even when reasons for assistance seeking or handovers at later stages in the process were included. Examples of cases where the normal procedures suggested that others had to be involved in the recovery process were cases where operators detected broken equipment that needed to be repaired, a task in which people from the maintenance and engineering departments specialize, and for which the operators may not even have the required expertise. Not having the required knowledge or expertise often also coincided with not having the necessary tools or equipment. Two additional reasons also involved not having the required resources, but in those cases, instead of knowledge or tools, the resources involved manpower or time. Another possible reason was that the need for a recovery step was already anticipated but that it could not be performed until that need would actually arise. An example of such a situation was the case where an operator, when reviewing the daily planning, recognized a ship that was scheduled to arrive later that day, and knew based on earlier experiences that an improvised type of connection would be needed for offloading this ship. Sometimes others would become involved purely by coincidence, for example when they were nearby when the event occurred. On other occasions, the only reason to involve others at some point in the process was to inform them about what was going on, to prevent potential additional problems. Reasons that did not fit into any of these categories were put in the category ‘other’.

Effect of the involvement of others

For almost all of the cases where during the process other people became involved, the people who decided to involve these others assured me that without the assistance of these others, it would not have been possible to complete all the necessary steps. This provided the answer to part *d)* of research question 5.

Further involvement of those who involved others

Even when a decision was made to involve others in the next step of a recovery process, this did not always mean that the person(s) previously involved stopped having any further role in the process. Table 15 provides an overview of what happened with regard to the further involvement of those who decided to involve other people. Recall that within one recovery process, there can be multiple moments at which those actually involved decide to involve others.

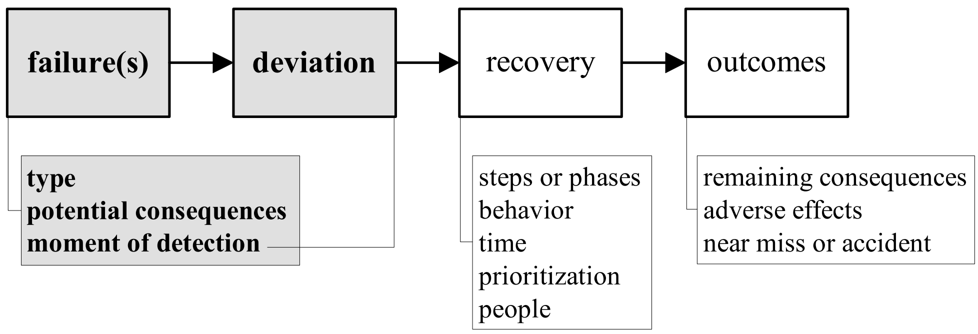
Table 15 Number of cases per specific further role of the persons who involved others, for each consecutive moment when other people became involved, for each of the studies

	number of cases											
	1st moment others are involved			**2nd moment others are involved**			**3rd moment others are involved**			**4th moment others are involved**		
further role	study 1	study 2	study 3	study 1	study 2	study 3	study 1	study 2	study 3	study 1	study 2	study 3
person stays involved for at least a while	32	15	21	12	2	10	9	1	7	1	0	2
person involved again at a later stage	3	1	4	1	1	1	0	0	2	0	0	0
no further involvement	13	2	6	20	4	11	7	0	3	1	1	3
total	*48*	*18*	*31*	*33*	*7*	*22*	*16*	*1*	*12*	*2*	*1*	*5*

From this table it is obvious that on most occasions when people decided to involve others in the next step(s) of the recovery process, they remained involved for at least a while. A complete handover was much more exceptional. In addition, on some occasions, when people had withdrawn their own involvement after handing over the process to others, they became involved again at a later stage, for example when a repaired part of the installation was taken back into operation.

Reasons for not involving others
There were also a number of cases in which no others were involved at all in the entire recovery process: as can be seen from table 13, two cases in study 1, 15 in study 2, and eight in study 3. The last section of research question 5 aimed at identifying the reasons for not involving any others. The reasons mentioned by the operators involved in these cases were: because the person who completed the recovery process alone was allowed to perform all the necessary steps, and responsible for performing the necessary steps, and, even if the other two conditions were not applicable, at least also able to perform the necessary steps alone.

3.3.6 Events leading up to recovery processes



The sixth research question linked recovery processes to the events leading up to them: *How are a) the failure types that precede a recovery process in a working environment, b) their potential consequences, c) the moment of detection (how long the deviation already existed, and performance stage when errors are detected), and d) the form the recovery process that follows takes with regard to steps, behavior, and people involved, prioritization and time spent, interrelated?*

Failure combinations
With regard to the failures preceding the recovery process, recall from section 2.3 that a deviation that needs to be recovered is often the result of not just one, but a combination of failures. This was also true for the cases collected during the three initial studies, as can be seen in table 16, containing an overview of the different failure combinations encountered in these cases and how often each combination played a role. In these combinations, T

stands for technical failures, H for human failures, O for organizational, and E for external to the organization. In the coding for failure combination, with regard to each failure type, the number of failures of this type was not taken into account, only the presence or absence. As a result, for example, a case categorized in the failure combination type T may involve multiple technical failures.

Table 16 Number of cases per specific failure combination, for each of the studies

failure combination	**number of cases in which failure combination played a role**		
	study 1	**study 2**	**study 3**
T (technical)	11	0	9
O (organizational)	5	0	2
H (human)	3	26	12
E (external)	1	0	1
TO	4	0	3
TH	3	2	3
TE	1	0	2
TOH	3	0	0
TOE	1	0	0
OH	17	3	5
HE	1	2	2
total	*50*	*33*	*39*

There were significant differences between the three studies with regard to the distribution of the cases across the different failure combinations, $\chi^2(20, N=122)=62.264, p<0.001$. An analysis of the adjusted standardized residuals revealed that study 1 contained significantly less cases involving human failures only, and significantly more cases involving organizational failures only, or combinations of organizational and human, or combinations of technical, organizational and human failures, than expected based on the marginal totals. Study 2 contained significantly more cases involving human failures only, and significantly less cases involving organizational failures only, or technical failures only, or combinations of organizational and human, or combinations of technical and organizational failures. In study 3 the distribution of the cases across the different failure combinations was not significantly different than expected. The most obvious difference, that all the cases in study 2 involved human error, most of them even exclusively, while the other studies included failure combinations without human error, was a result of the design of that study: people were asked to only report recovery from self-made errors.

Table 17 shows which relationship existed between failure combination and recovery process type (as discussed in subsection 3.3.1) that followed.

Table 17 Number of cases as a function of failure combination involved and consequent recovery process type

failure combination	**recovery process type**			
	D	**D→n*C**	**D→E→ {(k-1)*E & n*C}**	**D→C→ {k*E & (n-1)*E}**
T (technical)	0	1	11	8
O (organizational)	0	2	5	0
H (human)	1	22	15	3
E (external)	0	1	0	1
TO	0	3	4	0
TH	0	4	4	0
TE	0	0	1	2
TOH	0	1	2	0
TOE	0	1	0	0
OH	1	8	13	3
HE	0	1	4	0
total	*2*	*44*	*59*	*17*

For this table, observed values differed significantly from expected values given the marginal distributions for the dataset, $\chi^2(30, N=122)=42.656$, $p=0.010$, indicating that some recovery process types were more likely to follow specific failure combinations than others. An analysis of the adjusted standardized residuals revealed that while for the failure combinations O, E, TO, TH, TOH, TOE, OH, and HE none of the recovery process types occurred more or less frequently than expected based on row and column totals, for the failure combinations T, H, and TE there were significant differences from expected cell counts. For cases involving human failures only, significantly more recovery processes than expected were the type with only countermeasures following detection, and significantly less processes were the type where detection was followed by a combination of steps aimed at explanation and countermeasures, with the first explanation occurring immediately after detection. For cases involving technical failures only, significantly less recovery processes than expected were the type with only countermeasures following detection, and significantly more processes were the type where detection was followed by a combination of steps aimed at explanation and countermeasures, with the first step after detection belonging to the countermeasures phase. For cases involving a combination of technical and external failures also significantly more processes than expected were the type where detection was followed by a combination of steps aimed at explanation and countermeasures, with the first step after detection belonging to the countermeasures phase. The preference for specific recovery process types was tested again specifically for cases involving human failures only, collapsing the rows in the table involving all the other failure combinations, and differentiating only between cases with and without explanation. In this analysis ($\chi^2(1, N=122)=14.195$, $p<0.001$, using the asymptotic method and not the exact significance, as all of the necessary assumptions were met), the process types without explanation were more prevalent for human failures than process types with explanation. Where any other type of failure or failure combination had played a role, the opposite was true.

There was no significant relationship between failure combination and information processing behavior type (skill-, rule-, or knowledge-based) involved in the detection phase that followed, $\chi^2(20, N=122)=21.403$, $p=0.425$. No significant relationship existed either between failure combination and extent to which the detection was pre-planned by the organization, $\chi^2(30, N=122)=25.272$, $p=0.118$.

Table 18 shows the mean ranks for comparisons between the different encountered failure combinations of the total number of people involved in the consequent recovery process, the total amount of time between start and completion of the recovery process, and the total number of steps taken towards recovery. Mean ranks for Kruskal-Wallis tests are established based on rank ordering all cases from lowest to highest score on the ordinal variable involved: per value of the categorical variable involved in the analysis, the average of the corresponding ordinal variable rank numbers is calculated.

Table 18 Mean ranks for comparison between different failure combinations of total amount of people involved, total amount of time between recovery process start and completion, and total number of steps

failure combination	mean rank		
	total # people	**total time**	**total # steps**
T (technical)	88.68	73.18	88.20
O (organizational)	84.50	113.50	64.71
H (human)	38.07	37.50	44.28
E (external)	71.50	49.00	57.00
TO	82.07	97.07	79.36
TH	59.38	67.88	71.94
TE	101.17	98.00	101.17
TOH	89.67	96.17	87.00
TOE	36.00	108.00	9.00
OH	64.96	61.42	58.06
HE	30.40	32.20	40.10

Kruskal-Wallis tests indicated that there were significant differences between failure combinations with regard to number of people involved, total amount of time, and total number of steps between the different failure combinations, respectively $\chi^2(10, N=122)=46.519$, $p<0.001$, $\chi^2(10, N=122)= 56.565$, $p<0.001$, and $\chi^2(10, N=122)= 34.573$, $p<0.001$. I conducted follow-up Mann-Whitney tests to evaluate pairwise comparisons between the different failure combinations. To avoid type I error for multiple pairwise comparisons I adjusted the significance level for these tests to $\alpha=0.01$.

The results indicated that the total number of people involved in the recovery process only differed significantly between cases involving T versus cases involving H, or HE; between cases involving O versus cases involving H, or HE; between cases involving H versus cases involving TO, TE, TOH, or OH; and between cases involving TO versus cases involving HE. Generally speaking, the presence of human failures seemed to decrease the total number of people involved, whereas the presence of technical and organizational failures seemed to increase that number.

The results also indicated that the total amount of time between start and completion of the recovery process only differed significantly between cases involving T versus cases involving O, H, or HE as failure combination; between cases involving O versus cases involving H, TH, OH, or HE as failure combinations; between cases involving H versus cases involving TO, TE, TOH, or OH; and between cases involving TO versus cases involving OH, or HE. In general, the presence of human failures seemed to decrease the total amount of time, whereas the presence of technical and also organizational failures seemed to increase that amount.

Furthermore, the results indicated that the total number of steps involved in the recovery process only differed significantly between cases involving T versus cases involving H, or OH, and between cases involving H versus cases involving TE, with the combinations among those four where technical failures played a role consisting of a larger number of steps then where human failures played a role.

Potential consequences
With regard to the potential consequences if no recovery steps were to be taken, first of all I checked for differences between the studies. Table 19 contains an overview specifying how often which types of consequences at which severity level were applicable in each of the three studies. Note that within one case, several types of consequences may be applicable in combination.

Table 19 Number of cases per severity level and type of potential consequences if no recovery steps would have been taken, for each of the studies

type of potential consequence if no recovery	**number of cases**											
	study 1				**study 2**				**study 3**			
	no	minor	consi-derable	major	no	minor	consi-derable	major	no	minor	consi-derable	major
production / quality loss	35	11	4	0	19	9	4	1	21	10	5	3
delay	26	15	9	0	18	11	4	0	8	19	11	1
damage	31	16	3	0	25	7	1	0	19	7	11	2
injury / effects on health	27	8	11	4	24	3	5	1	13	6	13	7
environmen-tal spill	36	7	7	0	25	3	5	0	20	13	6	0

As it was difficult to compare the three studies at this level of detail, I calculated an overall score per case for potential consequences. To obtain this overall score, for each case I added up the scores on the different types of potential consequences. No consequences counted for 0, minor consequences counted for 1, considerable for 2, and major for 3. For each of the three studies, table 20 shows the distribution of the cases across the different overall scores on potential consequences.

Table 20 Number of cases per overall score on potential consequences, for each of the studies

overall score potential consequences	**number of cases**		
	study 1	**study 2**	**study 3**
0	0	2	0
1	6	10	0
2	14	9	6
3	15	3	10
4	8	6	1
5	5	2	10
6	0	1	5
7	1	0	2
8	1	0	0
9	0	0	4
10	0	0	1
total	*50*	*33*	*39*

A Kruskal-Wallis test revealed significant differences with regard to the overall score on potential consequences between the three studies, $\chi^2(2, N=122)=27.016$, $p<0.001$. Follow-up Mann-Whitney tests indicated that all three of the pairwise differences were significant, $z=-2.117$, $p=0.034$, $z=-3.873$, $p<0.001$, and $z=-4.701$, $p<0.001$ respectively for studies 1 and 2, 1 and 3, and 2 and 3. The average combined score on potential consequences was lowest for study 2, highest for study 3, and the average score for study 1 was in between these two. Another Kruskal-Wallis test indicated that the overall score on potential consequences also differed significantly between the four recovery process types, $\chi^2(3, N=122)=8.560$, $p=0.026$. Follow-up Mann-Whitney tests indicated that the recovery process types with only countermeasures after detection and process types with, after detection, both explanation and countermeasures, starting with explanation, differed significantly from the process types with, after detection, both explanation and countermeasures, starting with countermeasures. The former two involved lower scores on potential consequences than the latter. The overall score on potential consequences correlated positively with the total number of steps taken in the recovery process, $r_s(120)=0.331$, $p<0.001$, and with the total number of people involved, $r_s(120)=0.250$, $p=0.006$. In addition, as was indicated already in subsection 3.3.4, the overall score on potential consequences also correlated positively with the total amount of time passed between start and completion of the recovery process,

$r_s(120)=0.239$, $p=0.008$. Finally, a Kruskal-Wallis test indicated that there were significant differences with regard to the score on potential consequences between different combinations of failures that preceded the recovery process, $\chi^2(10, N=122)=23.511$, $p=0.003$. I conducted follow-up Mann-Whitney tests to evaluate pairwise comparisons between the different failure combinations. To avoid type I error for multiple pairwise comparisons I adjusted the alpha level for these tests to $\alpha=0.01$. The tests revealed that only the difference between cases involving human failures only and cases involving technical failures only was significant, the former cases had lower scores on potential consequences than the latter.

Another important characteristic of the potential consequences if no recovery steps were to be taken, besides type and severity that were combined in the overall score, is the time available until the earliest possible moment when these consequences can occur. This amount of time indicates how much time is available for the recovery steps, before it is too late for recovery. Table 21 shows for each of the three studies how the collected cases were distributed over the different time categories.

Table 21 Number (and percentage) of cases per category of amount of time until the earliest possible occurrence of potential consequences if no recovery steps were to be taken, for each of the studies

time until earliest possible occurrence of potential consequences	**number (percentage) of cases**					
	study 1		**study 2**		**study 3**	
up to 5 minutes	4	(8%)	8	(24.2%)	18	(46.2%)
between 5 minutes and 1 hour	7	(14%)	5	(15.2%)	8	(20.5%)
1 to 8 hours	5	(10%)	8	(24.2%)	3	(7.7%)
8 to 16 hours	0	(0%)	0	(0%)	1	(2.6%)
16 to 24 hours	0	(0%)	1	(3.0%)	0	(0%)
1 to 7 days	1	(2%)	2	(6.1%)	0	(0%)
1 to 4 weeks	0	(0%)	2	(6.1%)	0	(0%)
even longer than 4 weeks	1	(2%)	0	(0%)	0	(0%)
any time, hard to say	32	(64%)	7	(21.2%)	9	(23.1%)
total	*50*	*(100%)*	*33*	*(100%)*	*39*	*(100%)*

The cases where it was impossible to say what the earliest possible moment would be when consequences could occur, that is, when the actual occurrence of consequences also depended on other factors than simply time progressing, were not taken into account in the analyses in which this earliest possible occurrence moment played a role. An example of such a case would be the situation where a piece of equipment which is not always in use, is found to be broken, but whether this leads to negative consequences or not depends on it actually being needed while it is still broken. For this type of cases it was impossible to decide whether they belonged at the low end of the time categories or the high end. The people involved in recovery processes where there was such uncertainty generally took the required measures early on. A Kruskal-Wallis test indicated that there were significant differences between the three studies with regard to the time available until the earliest possible occurrence of the potential consequences, $\chi^2(2, N=74)=10.492$, $p=0.003$. Follow-up Mann-Whitney tests revealed that the differences between studies 1 and 3 and between studies 2 and 3 were significant, respectively $z=-2.684$, $p=0.005$, and $z=-2.873$, $p=0.003$. The time available until the earliest possible occurrence of the potential consequences if no recovery would take place was lower in study 3 than in studies 1 and 2. Another Kruskal-Wallis test revealed that the time available until the earliest possible occurrence of potential consequences did not differ significantly between the different recovery process types, $\chi^2(3, N=74)=4.629$, $p=0.177$. Furthermore, the time available until potential consequences did not correlate significantly with the total number of steps involved in the recovery process, $r_s(72)=-0.192$, $p=0.100$, nor with the total number of people involved, $r_s(72)=-0.012$, $p=0.916$. As was indicated already in subsection 3.3.4, the time available until the earliest possible occurrence of potential consequences also did not correlate significantly with the total amount of time passed between recovery process start and completion, $r_s(72)=0.0572$, $p=0.628$. The time available until the earliest possible occurrence of potential consequences did, however, correlate significantly with the overall score on potential consequences, $r_s(72)=-0.240$, $p=0.039$, which indicated that for the cases collected in the three studies, the more serious and widely noticeable potential consequences would also occur sooner if no recovery were to take place, than less serious consequences.

Moment of detection

Another characteristic of the events leading up to recovery processes is the moment of detection. One way to characterize this moment is by looking at the amount of time that has passed between the moment the failure occurred and its detection. Table 22 shows for each of the three studies how the collected cases were distributed over the different categories of elapsed time.

Table 22 Number (and percentage) of cases per category of amount of time elapsed between failure and detection, for each of the studies

time elapsed between failure and detection	number (percentage) of cases					
	study 1		**study 2**		**study 3**	
up to 5 minutes	9	(18%)	16	(48.5%)	14	(35.9%)
between 5 minutes and 1 hour	12	(24%)	7	(21.2%)	11	(28.2%)
1 to 8 hours	8	(16%)	6	(18.2%)	4	(10.3%)
8 to 16 hours	0	(0%)	2	(6.1%)	1	(2.6%)
16 to 24 hours	4	(8%)	1	(3.0%)	1	(2.6%)
1 to 7 days	11	(22%)	0	(0%)	7	(17.9%)
1 to 4 weeks	0	(0%)	0	(0%)	1	(2.6%)
even longer than 4 weeks	6	(12%)	1	(3.0%)	0	(0%)
total	*50*	*(100%)*	*33*	*(100%)*	*39*	*(100%)*

A Kruskal-Wallis test revealed significant differences between the three studies with regard to the time elapsed between failure and detection, $\chi^2(2, N=122)=13.025, p=0.001$. Follow-up Mann-Whitney tests indicated that the differences between studies 1 and 2 and between studies 1 and 3 were significant, $z=-3.451, p<0.001$, and $z=-2.315, p=0.019$, respectively. For study 1, the time elapsed between failure and detection was longer than for studies 2 and 3. Another Kruskal-Wallis test indicated that there were no significant differences between the different recovery process types with regard to the time elapsed between failure and detection, $\chi^2(3, N=122)=5.774, p=0.120$. In connection with this finding, there was no significant correlation either between the number of explanation steps and time elapsed between failure and detection, $r_s(120)=-0.017, p=0.858$, which could indicate that even if the time between failure and detection increased, explanation did not require more effort. An additional Kruskal-Wallis test indicated that there were significant differences between the different failure combinations (see table 16) with regard to the time elapsed between failure and detection, $\chi^2(10, N=122)=31.071, p<0.001$. I conducted follow-up Mann-Whitney tests to evaluate pairwise comparisons between the different failure combinations. To avoid type I error for multiple pairwise comparisons I adjusted the alpha level for these tests to $\alpha=0.01$. The results indicated that the time elapsed between failure and detection only differed significantly between cases involving T versus cases involving O, TO, TOH, or OH, the cases with technical failures were detected earlier; between cases involving O versus cases involving H, the cases with only organizational failures were detected later than cases with only human failures; and between cases involving H versus cases involving TO, TOH, or OH, whereby the human failures were detected earlier. Generally speaking, the presence of organizational failures seemed to increase the amount of time elapsed before a failure is detected. Furthermore, the time elapsed between failure and detection did not significantly correlate with the overall score on potential consequences, nor with the total number of steps involved in the recovery process, respectively $r_s(120)=-0.004$, $p=0.964$, and $r_s(120)=0.042$, $p=0.648$. On the other hand, the time elapsed between failure and detection correlated significantly and positively with the total number of people involved in the recovery process, $r_s(120)=0.232$, $p=0.009$. As could logically be expected, the time elapsed between failure and detection also correlated significantly and positively with the total amount of time elapsed between the start and completion of the recovery process, $r_s(120)=0.461$, $p<0.001$.

Another way to characterize the moment of detection is to look at the performance stage during which a failure is detected, that is, while planning the action, implementing it, or during the outcome stage. As discussed in section 2.3, these stages are referred to as planning, action or outcome stage, respectively. For each of the three studies, table 23 shows how the failure detection moments for the collected cases were distributed over the different performance stages. Since these performance stages concern human performance, only cases involving human failures in this performance were included in the following analyses, and cases in which not a single human failure was involved were excluded.

Table 23 Number of cases per performance stage during which a human failure is detected, for each of the studies

performance phase during which failure is detected	**number of cases**		
	study 1	**study 2**	**study 3**
planning phase	0	5	1
action phase	7	11	12
outcome phase	20	17	9
total	*27*	*33*	*22*

A Kruskal-Wallis test revealed significant differences between the three studies with regard to performance stage during which the failure was detected, $\chi^2(2, N=82)=6.222$, $p=0.042$. Follow-up Mann-Whitney tests indicated that the differences between studies 1 and 2 and between studies 1 and 3 were significant, $z=-2.041$, $p=0.047$, and $z=-2.392$, $p=0.016$, respectively. For study 1, the failure was detected generally in an earlier performance stage than in studies 2 and 3. Another Kruskal-Wallis test indicated that there were no significant differences between the different recovery process types with regard to the performance stage during which the failure was detected, $\chi^2(3, N=82)=0.827$, $p=0.842$. In connection with this finding, there was no significant correlation between the number of explanation steps and the performance stage during which the failure was detected, $r_s(80)=0.124$, $p=0.265$. An additional Kruskal-Wallis test indicated that there were no significant differences between the different failure combinations (see table 16, only those combinations involving human failures remained) with regard to the performance stage during which the failure was detected, $\chi^2(4, N=82)=8.841$, $p=0.057$. Furthermore, the performance stage during which the failure was detected did not significantly correlate with the overall score on potential consequences, $r_s(80)=0.046$, $p=0.684$, and neither with the total number of steps involved in the recovery process, $r_s(80)=0.108$, $p=0.346$, nor with the total number of people involved, $r_s(80)=0.134$, $p=0.224$. On the other hand, the performance stage during which the failure was detected correlated significantly and positively with the total amount of time elapsed between the start and completion of the recovery process, $r_s(80)=0.237$, $p=0.031$.

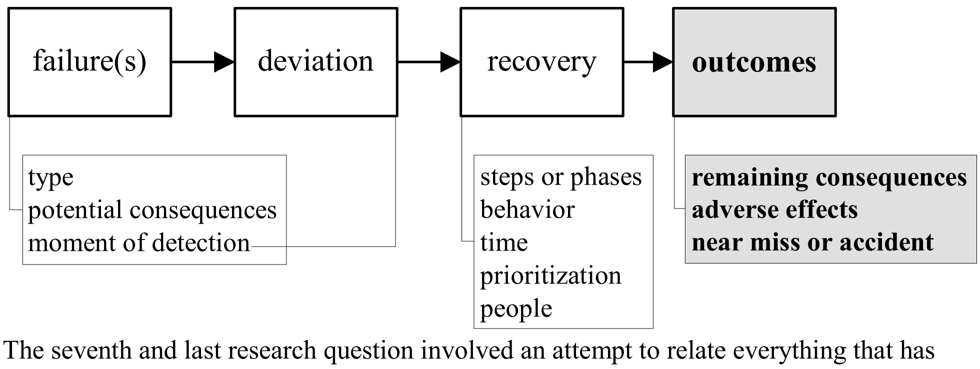
Human failures

Recall from section 2.3 that the only insights available from the literature on the relation between failures and recovery concerned recovery from human failures, i.e. errors. These suggested that the detection of skill-based errors would be easier than detection of rule-based and knowledge-based errors, would happen in an earlier performance stage, and less time would pass before the error was detected. In addition, as the nature of the error changes from skill-based to rule-based to knowledge-based, the amount of effort spent on explanation would increase. As in cases where human errors coincided with other types of failures it was not possible identify exactly which failure had most influence on how the subsequent recovery process turned out, it would not be justified to attribute everything to the human failures involved in these cases. This is why all failure combinations where other failure types than human failures played a role were excluded from the analyses performed to check if these expectations were confirmed in the data collected in studies 1, 2 and 3. Even though the number of cases in the dataset involving only human error was relatively small, namely 41, and four cases had to be dropped from this number because they involved combinations of different error types (e.g. skill- and rule-based, or rule- and knowledge-based), based on the remaining 37 cases it was still possible to test if the same findings applied here. Eight of these 37 cases involved skill-based errors, 24 cases involved rule-based errors, and five cases knowledge-based errors.

With regard to the relationship between error type and time elapsed between error and detection (also an indication of how easy it was to detect the error), a Kruskal-Wallis test revealed that there were no significant differences between the error types, $\chi^2(2, N=37)=0.526$, $p=0.781$. Another Kruskal-Wallis test indicated that there were no significant differences between the error types with regard to performance stage during which they were detected, $\chi^2(2, N=37)=0.008$, $p=1.000$. There were no significant differences between the error types with regard to behavior level involved in detection, $\chi^2(4, N=37)=5.139$, *p*=0.289, and with regard to the extent to which this behavior was pre-planned by the organization, $\chi^2(4, N=37)=4.662$, $p=0.451$.

On the other hand, significant differences were found between error types with regard to recovery process type that followed, $\chi^2(4, N=37)=9.661$, $p=0.041$. Analysis of the adjusted standardized residuals showed that knowledge-based errors involved significantly less recovery process types with only countermeasures after detection than expected based on the row and column totals, and significantly more recovery process types with detection followed by a combination, starting with explanation, of explanation and countermeasures. Corresponding with this finding, a Kruskal-Wallis test also revealed significant differences between the error types with regard to the number of steps in the recovery process that followed aimed at explanation, $\chi^2(2, N=37)=6.750$, $p=0.029$. Follow-up Mann-Whitney tests indicated that the differences between skill-based and knowledge-based errors and between rule-based and knowledge-based errors were significant, $z=-2.606$, $p=0.018$, and $z=-2.058$, $p=0.047$, respectively. The recovery processes for both skill- and rule-based errors involved less explanation steps (which could correspond with less explanation efforts) than the processes for knowledge-based errors. Finally, Kruskal-Wallis tests also indicated that there were no significant differences between the error types with regard to total number of people involved, total number of recovery process steps, and time elapsed between start and completion of the recovery process, $\chi^2(2, N=37)=5.084$, $p=0.077$, $\chi^2(2, N=37)=2.203$, $p=0.345$, and $\chi^2(2, N=37)=0.826$, $p=0.671$, respectively.

3.3.7 Relationship between preceding failures, recovery process characteristics and outcomes



The seventh and last research question involved an attempt to relate everything that has been discussed so far to the end result or outcomes of a recovery process: *What is the relationship of a) the failure types that precede a recovery process in a working environment, b) their potential consequences, c) the moment of detection (how long the* *deviation already existed, and performance stage when errors are detected), and d) the form the recovery process that follows takes with regard to steps, behavior, and people involved, prioritization and time spent, with recovery process outcomes?*

Recall that in section 2.4 two variables were proposed to measure the outcomes of a recovery process: the adverse effects of well-intended recovery steps, and the remaining consequences after all recovery steps have been completed. Among all the cases collected in the three studies, there was only one case where unintended adverse effects took place: an operator burnt his index finger and thumb in an attempt to refill cooling water without having put on proper personal protective equipment. Of course, cases where unintended adverse effects of recovery actions were really bad would qualify as accidents, for which different reporting systems applied than the near miss reporting system that was used as a basis for the first study. In the second and third study did no existing reporting channels were used, and in the instruction given to the participants it was made clear that both examples of successful recovery and of less successful recovery were welcome. As such, it was possible to report recovery efforts with adverse effects in these studies, but still, participants may have preferred to share their more successful attempts over anything with actual adverse effects. While the lack of adverse effects of recovery efforts is a positive sign of people's abilities in recovery, it also meant that this variable could not be used to distinguish between processes, and the remaining consequences were left as the only usable measure of the recovery process outcomes.

With regard to these remaining consequences, similar as for the potential consequences in subsection 3.3.6, several types of consequences can be distinguished, and per type several severity levels. Table 24 contains an overview specifying how often which types of remaining consequences at which severity level were found in each of the three studies. Here, too, within one case several types of consequences may be applicable in combination.

Table 24 Number of cases per severity level and type of remaining consequences, for each of the studies

type of remaining consequence	number of cases								
	study 1			study 2			study 3		
	no	minor	consi-derable	no	minor	consi-derable	no	minor	consi-derable
production / quality loss	50	0	0	32	1	0	36	3	0
delay	44	6	0	30	3	0	30	8	1
damage	50	0	0	33	0	0	37	2	0
injury / effects on health	49	1	0	33	0	0	35	4	0
environmental spill	46	4	0	33	0	0	38	1	0
repair or correction costs	30	20	0	32	1	0	26	6	7
prolonged exposure to hazard until final correction	36	14	0	33	0	0	33	6	0

For each case, again similar to how this was done for potential consequences, an overall score on remaining consequences was calculated, by adding up the scores on the different types of remaining consequences per case. No consequences counted for 0, minor consequences counted for 1, and considerable for 2. For each of the three studies, table 25 shows the distribution of the cases across the different overall scores on remaining consequences.

Table 25 Number of cases per overall score on remaining consequences, for each of the studies

overall score remaining consequences	number of cases		
	study 1	study 2	study 3
0	16	29	18
1	24	3	6
2	9	1	7
3	1	0	7
4	0	0	0
5	0	0	1
total	*50*	*33*	*39*

A Kruskal-Wallis test revealed significant differences between the three studies with regard to the overall score on remaining consequences, $\chi^2(2, N=122)=22.644, p<0.001$. Follow-up Mann-Whitney test indicated that the differences between studies 1 and 2 and between studies 2 and 3 were significant, $z=-4.818, p<0.001$, and $z=-3.895, p<0.001$, respectively. The cases in study 2 had a lower score on remaining consequences than the cases in studies 1 and 3.

Relationship between characteristics of events preceding recovery and outcomes
A Kruskal-Wallis test indicated that there were significant differences with regard to the score on remaining consequences between different combinations of failures that preceded the recovery process, $\chi^2(10, N=122)=37.948, p<0.001$. I conducted follow-up Mann-Whitney tests to evaluate pairwise comparisons between the different failure combinations. To avoid type I error for multiple pairwise comparisons I adjusted the significance level for these tests to $\alpha=0.01$. The tests revealed that there were significant differences between cases involving human failures only (failure combination H) and cases involving failure combinations T, O, TO, TH, TE, TOH, or OH, among these the cases involving human failures only had a lower remaining consequences score. In addition, cases involving failure combination OH had a significantly lower score on remaining consequences than cases involving failure combination TO. A significant, positive correlation was found between the score on potential consequences and the score on remaining consequences, $r_s(120)=0.405, p<0.001$. There was no significant correlation between the time available until the earliest possible occurrence of the potential consequences and the score on remaining consequences, $r_s(72)=-0.018, p=0.878$ (to assess this relationship, the cases where it was impossible to say when potential consequences could occur were left outside the analysis). There was a significant, positive correlation between time elapsed before a failure is detected and the score on remaining consequences, $r_s(120)= 0.284, p=0.002$. The correlation between performance stage during which a failure is detected and the score on remaining consequences was close to being significant, $r_s(80)= 0.217, p=0.051$ (for this analysis, only cases involving recovery from failure combinations that included human failure were taken into account, for the same reasons as mentioned in subsection 3.3.6).

Relationship between recovery process characteristics and outcomes
A Kruskal-Wallis test indicated that there were no significant differences with regard to score on remaining consequences between the four different recovery process types, $\chi^2(3, N=122)=2.528, p=0.488$. Even when only a distinction was made between processes with or without explanation, or when the number of steps aimed at explanation was taken into account, the analyses still did not produce any significant results, respectively $\chi^2(4, N=122)=4.763, p=0.302$, and $r_s(120)= 0.058, p=0.527$. There was a significant, positive correlation between total number of recovery steps and score on remaining consequences,

$r_s(120) = 0.365$, $p<0.001$. A significant, positive correlation also existed between time elapsed from start to completion of the recovery process and score on remaining consequences, $r_s(120) = 0.630$, $p<0.001$. Furthermore, a significant and positive correlation was found between total number of people involved in the recovery process and score on remaining consequences, $r_s(120) = 0.437$, $p<0.001$. Finally, a significant, positive correlation was found between the extent to which tasks other than recovery were present that also required attention and score on remaining consequences, too, $r_s(120) = 0.487$, $p<0.001$.

3.4 Discussion and conclusions based on initial studies

3.4.1 Implications of the findings

As for the *first research question*, concerning recovery steps or phases and the sequence or order in which they are performed: the proposed descriptive recovery process model was verified, both with regard to steps or phases involved and possible transitions between them. In line with the proposed phase model, a recovery process, per definition, always started with detection, and after that, any number and combination of steps aimed at explanation and countermeasures were possible. This finding also demonstrated that linear models, such as those proposed by, for example, Johannsen (1988), Zapf & Reason (1994), and Bove & Andersen (2001), are a rather simplified way to describe recovery processes. Where researchers like Bagnara et al. (1988), Van der Schaaf (1988), and Kontogiannis (1999) already pointed out that the explanation phase does not necessarily have to occur, the findings from the initial studies took this a step further and did not only demonstrate that the phases that follow detection do not necessarily have to occur, but also that recurrences or iterations of these phases are possible. The number and timing of explanation steps taking place in the recovery process turned out to be a useful way to distinguish between processes. In addition to the verification of the high-level phase model, based on the specific goals of separate recovery steps, further detail was added to the model, similar to how this was done by, for example, Norman (1984), who distinguished separate stages and levels of activity in human-machine interaction, and Newell and Simon (1972), who distinguished separate episodes in problem solving, each with their own subgoal. In the explanation phase two sub-phases were distinguished, definition of the problem, and identification of the causes, and within the countermeasures phase four sub-phases, stabilization, mitigation, temporary correction, and permanent correction. Again, not all of these sub-phases needed to occur within any given recovery process, and iterations of sub-phases were possible too. The newly identified sub-phases proved to be useful in establishing the reasons why certain recovery steps were performed at certain stages in the process. In cases involving steps aimed at stabilization or mitigation, these generally occurred early on in the process, as expected, to buy time for further steps. Furthermore, most processes ended with a countermeasures phase, except for a few cases where after all the countermeasures were completed the organization's management felt that additional lessons could be learnt from a causal analysis, and such an analysis had not occurred up to that moment. Processes where some effort was spent on explanation involved more steps than processes where this was not the case, no matter whether the first explanation efforts took place immediately after detection or only at a later stage. In the latter case, there was usually a good reason to implement some countermeasures first before any effort was spent on explanation, that is, to bring the situation under control first. A larger number of steps, however, did not necessarily mean that this increase was purely the result of additional explanation steps, it equally involved additional countermeasures steps. When no substantial efforts were spent on explanation, this did not necessarily mean that no lessons were learnt from the event. In many of such cases at the moment of detection both the extent and the causes of the problem were already obvious without further analysis. The observation that there were hardly any cases where no countermeasures were taken at all was a good indication of the importance given to recovery by the people in the organizations that participated in the studies.

> *In summary:*
> *Recovery processes were shown always to start with the detection of a deviation, as per definition, and after that, any number and combination of steps aimed at explanation and countermeasures were possible, ranging from no such steps occurring at all to multiple recurrences of either one, most often ending with a countermeasures phase. Several sub-goals were identified within the explanation phase (definition of the problem and identification of the causes) and the countermeasures phase (stabilization, mitigation, temporary correction, and permanent correction).*

Regarding the *second research question*, concerning the extent to which the recovery steps were pre-planned (cf. Mintzberg, 1976; Allwood & Montgomery, 1982) and the information processing behavior levels involved, the research showed the distinction between skill-based, rule-based, and knowledge-based behavior levels, widely used in the study of errors (e.g. Rasmussen, 1986; Reason, 1990), and whereby each level requires a different kind of support, to apply to recovery steps as well. The difficulty in coding for this behavior was that one recovery step could involve performance at several behavior levels, for example steps where a specific goal and a plan for reaching it were established at knowledge-based level, but once this was done many of corresponding actions could be performed at rule-based level. In such cases, the entire process step was coded as involving the highest of the applicable behavior levels. The research demonstrated the importance of ad-hoc steps, which was confirmed by the relatively large contributions of knowledge-based behavior, and which was especially unexpected for the detection phase. An explanation for the ad-hoc steps found in the detection phase may be that the operators had a very detailed and accurate mental model of the plant and processes, which made it easier for them to detect deviations from how things should be, even if they were not specifically told to look for certain things via procedures, or warned about situations via alarms. The important role of ad-hoc and knowledge-based behavior was expected for the explanation phase, and, as expected, turned out to be more dominant in this phase than in the other phases. This indicated that deviations can have characteristics that are very specific to the situation, so 100% pre-planning is hardly ever possible for their explanation, and neither is rule-based behavior, nor, especially, skill-based behavior. Therefore, next to providing procedures or work routine specifying where to look for what, organizations should also provide their operators with a substantial amount of knowledge about the process they are controlling, which can help them to identify and explain situations where procedures do not offer enough assistance. Even though the explanation phases involved mainly unplanned steps, there were a few cases where the explanation was supported by pre-planned provisions, mostly in the form of procedures, although there were also cases where the process control panel played a role. The preparedness of the organizations for problem situations was most visible in the countermeasures phase, even though available provisions in the form of procedures or tools regularly also still had to be tailored or fine-tuned to specific situations. For the countermeasures phases, one of the main reasons planned steps were commonly found here may be because quick recovery was ensured. That is, to select the appropriate countermeasures not much time was needed once one knew what the problem was; they did not need to be generated from scratch. Occasionally, in situations involving unusual, rare or unforeseen problems, unplanned countermeasures were implemented. In those cases where unplanned countermeasures were implemented, it was either an instinctive reaction, a reaction based on common sense, or a creative solution to gain time for planned countermeasures that would normally take a while to complete. These findings underlined the importance for organizations to invest in their employees' plant- and process in-depth knowledge, i.e. the employees' mental model, in order to support all recovery phases, to prepare them for the unexpected. Such knowledge proved to be beneficial in many of the analyzed recovery steps, where people were able to determine what to do without specifically being instructed to do so.

> *In summary:*
> *The research demonstrated the important role of ad-hoc and knowledge-based behavior in recovery processes, that is, of people's creative thinking abilities. Most pre-planned steps were found in the countermeasures phase.*

With regard to the *third research question*, concerning time spent on separate recovery steps and the entire process, a key finding was that steps taken at early stages in the process could usually be completed within very little time. That detection hardly ever took more than five minutes should be considered in light of the definitions of the process phases: all one needs to detect is that something is wrong, that there is a deviation from the normal, expected situation. Any additional search for more information about the deviation was considered as part of the explanation phase, such steps were either aimed at finding out more about the extent of the deviation and how it had manifested itself, or at finding what caused the deviation, i.e. the failure(s) that played a role. Other steps typically performed in early stages were those aimed at stabilization and mitigation. What may have helped their speedy completion was that they often involved skill-based, so well-rehearsed, almost automatic behavior. Generally speaking, this speed was essential to prevent things from getting further out of hand. Another important finding was that pre-planned countermeasures in general took less time to complete than ad-hoc countermeasures or steps where both types of behavior were combined. This indicated that for deviations where a fast recovery is essential, organizations should focus on providing pre-planned support for the necessary countermeasures. Furthermore, the more conscious reasoning a step involved, the more time was spent on the step. The finding that only a very small percentage of all the steps involved in the collected cases took more than one hour to complete is in itself not such a good indication of the recovery performance of the organizations participating in the studies. After all, not only when a step involved a different goal than what people were striving for before, but also every time the process was put on hold for a while, or when others became involved, whatever was done after that was considered and coded as a separate step. A better indication of an organization's recovery performance with regard to time may be to look at the total amount of time elapsed between the start and the completion of the recovery process, even though this amount can include some 'idle' time when no steps towards recovery are taken. With regard to this 'idle' time, organizations should make sure that the potential hazards associated with a deviation are contained within as little time as possible, early on in the process, before other tasks are allowed to delay the progress of the recovery process. The total amount of time elapsed between process start and completion, not surprisingly, correlated positively with the number of recovery steps taken, but did not differ significantly between processes with or without explanation. While it might be tempting to think that this latter finding indicated that it does not make a difference with regard to overall amount of time elapsed between process start and completion whether any effort is spent on explanation or not, care should be taken not to jump to such a conclusion. The consideration whether explanation efforts are needed should really depend on the situation involved. If an explanation is needed and people start attempting countermeasures without one, chances are that they will not find the right countermeasures and just waste valuable time trying until they arrive at a proper explanation of the deviation first.

> *In summary:*
> *Pre-planning of countermeasures was found to significantly speed up these steps. The early steps in recovery processes were shown to be completed the fastest, their goal often being to bring things back under control. The overall time between recovery process start and completion did not only depend on time spent on the recovery steps, but also on the time elapsed in between those steps, which underlined the importance of getting the potential hazards associated with a deviation contained early on in the process, before other tasks are allowed to delay the progress of the recovery process.*

Regarding the *fourth research question*, concerning prioritization of steps when recovery is needed, the only measure available in the dataset that could be used as an indication of the priority given to recovery was the total amount of time passed between the detection of a deviation and completion of the recovery process. This measure was probably not the best indication of the priority given to recovery steps. It may be better instead to look at how long people waited before starting the first recovery step after detection (cf. Waller, 1999). Unfortunately, the dataset did not contain any usable measures of the amount of time passed between detection and the first recovery step that followed. Using the total amount of time elapsed between recovery process start and completion as an indirect measure of priority enabled me to draw at least a few conclusions. The research showed that the processes tended to take longer in situations where a larger amount of other tasks also required urgent attention. This made sense, as in those situations not all available human resources can be deployed in recovery. As the amount of tasks other than recovery also requiring attention increased, the total number of separate recovery steps increased as well, which could indicate that people in such situations try to distribute their attention over the different tasks, changing their focus back and forth between recovery and other tasks, possibly even splitting up tasks into smaller fragments which they perform one at the time. The research also indicated that recovery processes tended to take longer in situations where more, and more severe negative consequences were possible if no recovery were to take place. This finding was difficult to interpret: it could be that it takes more work to stop diverse and more severe potential consequences from occurring, or that the people involved deemed such situations worthy of more time being spent on them, or possibly both, or an entirely different explanation may apply. Finally, the time available until the earliest possible occurrence of the negative consequences if no recovery would take place did not significantly affect the total amount of time between recovery process start and completion. Of course this factor may have influenced how long people waited before starting the first recovery step after detection, but this, as was pointed out above, could not be tested. The effects of the time and means necessary for recovery were not measured in a manner suitable for structured analysis, either. With all this information lacking, no attempts were made to look at the effects of all the factors influencing prioritization in combination, i.e. taking all factors into account at once instead of one by one. A final remark is that with regard to the factors that appear to play a role in prioritization of recovery steps, organizations may want to make sure that their employees are able to make correct assessments of the values of these factors in any given situation where recovery is needed. This way, incorrect prioritization based on the wrong assessments can be avoided.

> *In summary:*
> *The presence of other urgent and important tasks was shown to influence prioritization, in that they led to an increase in the total amount of time that elapsed between recovery process start and completion. The dataset did not contain usable information for a possibly better measure of the priority given to recovery tasks, that is, how soon after detection the next recovery step started, and neither on the estimates people made at the start of the process, of means and time needed for recovery.*

As for the *fifth research question*, concerning the people who play a role in recovery, the research demonstrated that most processes involved steps from more than just one person. In most cases, even in several cases involving self-made errors, detection was done by others than those involved in the preceding failures. In most of the recovery processes in all of the three studies, at some point additional people became involved. Sometimes this happened as early as during the detection phase, for example when people were working together, but otherwise at later stages. Most often, they were needed because of their qualifications, expertise, equipment they had, or the additional capacity they could offer, or for a combination of those reasons. Without their assistance, as the people involved claimed when they were asked about this, in most cases it would not have been possible to complete the recovery processes. These findings indicated that the involvement of others indeed was not only important in the detection phase, as was already pointed out by, for example, Hutchins (1996), Doireau et al. (1997), and Wioland & Amalberti (1998), but also in later stages of the process. In the light of this finding, organizations may want to make sure that access to and communication with other people who may be needed in recovery processes is made as easy as possible. Longer recovery processes with regard to number of steps and total amount of time between start and completion tended to involve more people than shorter processes. An explanation for this relationship could be that as time passes, eventually new personnel will come on duty, who then can continue a process that was started by their co-workers in a previous shift. Or the relationship could be the other way around, the recovery process may have to wait for new people to arrive, for example because their skills and expertise are needed, and therefore take longer. The people who decided to involve others most often remained involved for a while, and sometimes they also became involved again at a later stage, which served as an indication of their commitment to the progress of the recovery process.

> *In summary:*
> *The involvement of others was found to be not only important for the detection of deviations, but also for later steps in the recovery process, because of their qualifications, expertise, equipment they had, or the additional capacity they could offer, or a combination of those reasons.*

Regarding the *sixth research question*, concerning the relations between preceding failures, their potential consequences, the moment of detection, and characteristics of the recovery process, there were several significant findings. Recovery processes in cases involving human failures only most often did not involve any steps aimed at explanation, but as soon as other failure types played a role, most often the recovery process included at least one explanation phase. Furthermore, it was not possible to establish which of the failures, if a combination of different types had played a role, had the largest influence on the steps involved in recovery. For the sake of internal validity, I felt it would be best to do justice to precisely what happened, and take the entire combination into account when dividing the processes over different categories of preceding failures. When technical failures played a role not in combination with human failures, recovery involved a larger number of steps then when human failures played a role, not in combination with technical. The presence of human failures seemed to decrease the total number of people involved, whereas the presence of technical and also organizational failures seemed to increase that number. The same applied for the total amount of time between process start and completion. These three variables, total number of steps, people, and total amount of time, also correlated positively among each other.
The score on potential consequences did not differ much between different combinations of failure types, only the difference between cases involving only technical failures and cases involving only human failures was significant; whereby human failures corresponded with lower scores on potential consequences than technical failures. For high scores on potential consequences, i.e. more and/or more severe potential consequences, the preferred recovery process type was that where some countermeasures were taken immediately after detection, before the first explanation step took place. This could be an indication that for severe and/or a multitude of potential consequences the need is felt to at least ‘act a little’ first, before spending time on explanation, a phase which does not lead to directly noticeable results. There were positive correlations between score on potential consequences and the

total number of steps involved in recovery, the total number of people involved, and the total amount of time between process start and completion.

The time available until the earliest possible occurrence of potential consequences if no recovery were to take place did not differ between different types of processes, different numbers of steps involved in the process, different numbers of people involved, and different amounts of time passed between process start and completion. From these findings it may seem that this time until the earliest possible occurrence of consequences did not have any impact on the recovery process that followed. However, as was said before, the time that passed between detection and the first recovery step taken after this detection was not measured, but could very well have been influenced by the time available until negative consequences could occur. The negative correlation of time available until the earliest possible occurrence of negative consequences with the overall score on potential consequences indicated that the more severe consequences would also occur sooner. This does not necessarily mean that potential consequences that are only likely to occur after longer periods of time really are not as severe as those that can occur at a more foreseeable term. This finding may very well have been a consequence of the difficulty to predict events in a relatively remote future.

With regard to the moment of detection, the time elapsed between the occurrence of the failures and detection did not differ significantly between different types of recovery process, nor between different amounts of explanation taking place. This indicated that even if a failure occurred longer ago, this does not necessarily mean that it takes more effort to find an explanation, nor that there is more need for an explanation. Similarly, longer amounts of time between failure and detection did not lead to an increase in total number of steps involved in recovery. There was no difference with regard to potential consequences between the different amounts of time after which failures were detected, which indicated that people were not only alert for potential disasters, but equally alert for potentially less serious events as well. In cases involving organizational failures, more time passed before they were detected than in cases without organizational failures, which demonstrated the latent nature of such failure types. Latent failures (e.g. Reason, 1990) are usually present in the background, hardly noticeable, but they have the potential to lead to negative consequences. Latent failures may become manifest if this potential is triggered by an initiating event. In cases where more time had passed between failure and detection, more people were involved in the recovery process, which could be an indication that joined efforts were more needed in such cases, possibly because the deviation had become more widespread, affecting more people's work by then. In cases where more time had passed between failure and detection, the recovery process also took a longer amount of time, which could indicate that in such cases recovery becomes more difficult, and a speedy recovery process is no longer very likely. With regard to the performance stage in which the failures were detected, the same findings and interpretations apply as those for the time

elapsed between failures and detection, whereby an early performance stage corresponds with fast detection, except for two differences. The first difference was that there were no significant differences between failure combinations with regard to performance stage during which the failures were detected. The other difference was that the number of people involved did not differ significantly between different performance stages during which a failure was detected. Possibly these relationships were suppressed by the presence of human failures in all of the cases that were taken into account for the corresponding analyses (for cases not involving human failures it does not make sense to look for a stage in human performance during which the failure occurred). Thus, only the positive correlation of the total amount of time between process start and completion with the performance stage during which a failure was detected was significant.
Based on the moment in time when everything occurs or becomes known, the identified relations between preceding failure combinations, potential consequences, time available until their earliest possible occurrence, moment of detection, total amount of time spent on recovery, types and number of steps, and number of people involved, can be combined into a larger pattern in which tentative conclusions can be drawn regarding causality.

> *In summary:*
> *As compared to human failures, if technical and organizational failures had played a role in causing a deviation, this tended to increase both the number of explanation steps and the number of people involved in the recovery process, and these processes eventually also took longer before they were completed. There was not much difference in potential consequences between different types and combinations of failures. For deviations involving more and more severe potential consequences, there was a tendency to first implement some countermeasures before time was taken for explanation. Deviations involving more, and more severe potential consequences tended to lead to longer recovery processes, involving more steps and more people. If organizational failures played a role in causing a deviation, they tended to delay the detection moment, demonstrating their often latent nature. Late detection tended to lead to longer processes involving more people. Detection in a later performance stage also tended to lead to longer recovery processes.*

While for the analysis of the relationship between other failure types than exclusively human failures and the recovery process that followed there were no existing insights on which predictions could be based, for cases involving only *human failures* (i.e. errors) such insights were available. Some of the predictions for the analyses specifically focusing on recovery from only human failures were confirmed, but no support was found for a few other predictions. The lack of support for the prediction that the detection of skill-based errors would be easier than the detection of rule- and knowledge-based errors (based on

Embrey & Lucas, 1988; Zapf et al, 1994) could simply be the result of the small number of cases in my dataset involving only human failures. Among these cases, especially the numbers of skill-based errors and knowledge-based errors were very small, eight and five respectively. Compared to how the recovery from skill- or rule-based errors was distributed over the different recovery process types, recovery from knowledge-based errors involved significantly fewer processes with no explanation at all, and significantly more processes with both explanation and countermeasures, with explanation immediately following detection. This finding was in line with findings from Bagnara et al. (1988) and Sellen (1994). Furthermore, I also found that recovery from either skill-based or rule-based errors involved significantly fewer steps aimed at explanation than recovery from knowledge-based errors, which was again in line with the predictions based on these researchers' findings. These findings indicated that in cases involving knowledge-based errors, the extent of the resulting deviation, its causes, and what to do is hardly ever immediately obvious, instead, effort needs to be spent on finding an explanation, before further recovery is possible.

> *In summary:*
> *No support was found for other researchers' findings that skill-based errors are easier to detect than rule- and knowledge-based errors, possibly as a result of the small number of cases involving only human failures. In line with other researchers' findings, recovery from knowledge-based errors was found to involve more efforts aimed at explanation than recovery from skill- or rule-based errors.*

With regard to the *seventh research question*, concerning the relationship of preceding failures, their potential consequences, the moment of detection, and characteristics of the recovery process with recovery process outcomes, two outcome measures were proposed: the remaining consequences, and possible adverse effects of the recovery steps. The lack of cases in the dataset involving adverse effects from well-intended recovery steps was a positive sign of people's abilities in recovery, but also meant one variable less that could be used as a measure of the recovery process outcomes.

There were no significant differences in scores on remaining consequences between different types of recovery processes or between processes with or without explanation, which indicated that efforts aimed at explanation are not always needed to bring a recovery process to a good ending, and that the occurrence of one or more explanation phases and the moment at which they occur are no guarantee for successfully avoiding any of the distinguished types of remaining consequences either. As was pointed out already in a previous paragraph, the consideration whether explanation efforts are needed should really depend on the situation involved. If an explanation is needed and people start attempting countermeasures without one, chances are that they will not find the right countermeasures, and therefore end up with more or more severe remaining consequences than when they would have established a proper explanation of the deviation first. There was also no relation between time available until the earliest possible occurrence of negative consequences and the remaining consequences, which was actually a little surprising. Maybe the effects of the time available until consequences can occur are only noticeable at the beginning of the recovery process.

The results demonstrated that the longer the time was that elapsed between failures and detection, the higher was also the score on remaining consequences. Late detection obviously made a complete and successful recovery more difficult. Furthermore, the more people the processes involved, the more steps the processes involved, and the more time passed between process start and completion, the higher were the scores on remaining consequences. These findings can be connected with findings that were discussed in previous paragraphs, regarding the correlations among these variables, the relation between the moment of detection and total amount of time passed from recovery process start to completion, and the relation between the moment of detection and total number of people involved. A pattern then emerges, based on the order in time in which everything takes place or becomes known:

> *In summary:*
> *Late detection leads to longer processes in which more people are involved, such processes also involve more steps, and as result of all of these things combined, more and/or more severe consequences remain at the end.*

In the light of this conclusion, organizations may want to look for ways to reduce detection time as much as possible to improve their chances of successful recovery. An additional, significant finding was a positive correlation between scores on potential consequences and scores on remaining consequences. This could indicate that if the consequences can be really serious it may be difficult to avoid all of them, whereas in the case of less severe potential consequences this can be achieved more easily. Finally, cases involving only human failures had less remaining consequences than cases involving technical and/or organizational failures, no matter whether this was in combination with human failures or not. These findings can be linked with the other findings regarding recovery processes in cases involving high potential consequences scores, and regarding recovery processes following specifically human failures only versus processes following other types of failures and combinations of failures. Based on the order in which everything takes place or becomes known, the following pattern then emerges:

In summary:
Cases involving technical and/or organizational failures and/or many or severe potential consequences lead to longer processes, involving more steps and people than cases involving only human failures and few or minor potential consequences, and as a result of all these things combined, more and/or more severe consequences remain at the end.

While there may clearly be some room for improvement regarding the ability to recover from failure combinations involving technical and/or organizational failures, this does not necessarily mean that recovery from only human failures does not deserve any additional attention. The potential consequences if no recovery steps were to be taken can be very serious also in cases involving only human failures, so organizations should take care not to lose their current expertise in this area.

Table 26 contains an overview of all the statistically significant relations that were found between the variables characterizing the recovery process. Where applicable, an indication is given whether the relation concerns a negative or positive correlation.

Table 26 Relations between the recovery process variables

variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) time available until consequences can occur		-									
(2) potential consequences	-		*				+	+		+	+
(3) failure combination		*		*		*	*	*		*	*
(4) time between failure and detection			*					+		+	+
(5) performance stage when failure detected								+		+	+
(6) recovery process type / explanation			*				*	*			
(7) # steps		+	*			*		+	+	+	+
(8) # people		+	*	+	+	*	+		+	+	+
(9) presence of other tasks							+	+		+	+
(10) total time between process start and completion		+	*	+	+		+	+	+		+
(11) remaining consequences		+	*	+	+		+	+	+	+	

+ significant (p<0.05) positive correlation
- significant (p<0.05) negative correlation
* significant (p<0.05) relationship involving at least one categorical variable

3.4.2 Differences between the three studies

While I acknowledge the potential criticism that the differences that were found between the three studies could to some extent be a result of the different data collection methods that were used, at least care was taken to collect exactly the same information for every case in each of the studies, so they could be combined into one dataset on which the analyses were performed. The identified differences were never used to draw conclusions with regard to the participating organizations' recovery performance; that was not an objective of the studies. The reason for using the different data collection methods was merely the result of the objective to collect data on as many different types of cases as possible, to create a dataset representative of all these types. The dataset was not representative of the actual frequencies of occurrence of the different types within the participating organizations, but no conclusions regarding such frequencies were drawn from the data either. Research (Van der Schaaf & Kanse, 2004) has shown that even in organizations that have a well-functioning near miss reporting system, for various reasons the reporting frequencies of most types of events are not representative of the actual occurrence frequencies of such events. The dataset was only used to look for patterns and relationships within these different types of cases, i.e. how often did a certain value of one variable occur together with a certain value of another. Furthermore, there are much more plausible explanations for the many differences between the studies than simply the different data collection approach, as discussed below.

The main difference between the studies concerned the preceding failures. This difference was only significant between study 1 and 2; study 2 involved more human failures and less technical and organizational, while the exact opposite was true for study 1. This difference was intentional, the idea behind the second study was that of theoretical replication, I wanted to know if self-made human failures were followed by different processes and outcomes, and if so, why. The mixture of failure combinations in study 3 was more balanced.

Cases in study 2 involved fewer steps, fewer people, and less time between process start and completion than cases in the other two studies, findings that correspond with the findings for recovery from human failures only.

The finding that cases in study 2 involved fewer other tasks requiring immediate attention may have been a coincidence, or people may have felt a responsibility towards recovering their own errors, causing them to perceive recovery as more urgent even if their attention was also needed for other tasks; in such cases they may have rated the priority of those other tasks at a lower level.

The average overall score on potential consequences was lowest for cases in study 2, a little higher for study 1, and highest for study 3. In study 2, I may have lowered the reporting threshold with regard to potential seriousness by asking for a specific type of event that normally did not get reported. Also, the time available until the earliest possible occurrence of negative consequences was lower for cases in study 3 than for cases in studies 1 and 2. Parts of these findings correspond with the findings regarding the relation between failure types and potential consequences: when only human failures were involved the score on potential consequences was lower. An additional explanation, however, lies in the difference between the two organizations. While reporting and analyzing near misses was a common practice in the organization where the first two studies were performed, this was much less so in the organization participating in the third study. This may have made the reporting threshold there with regard to potential seriousness a little higher. Many of the people I interviewed in the third study did not have a firm grasp of what constituted a near miss before I started talking to them about such events. They tended to tell me about slightly more serious events with regard to what could have happened and how soon that could have been. They also had to dig into their memories to find the sort of events I was looking for, which was in fact indeed a result of the data collection method. It may have been easier to recall the potentially more serious cases. The chance that what people told me about events involving recovery efforts was distorted by having to base their story on their memories was kept to a minimum. The majority of the events had happened in the days just before they told me about them, so their memories of what happened were still fresh. Furthermore, the introduction I had given to the people in this plant regarding my research project, made them more alert on remembering the type of events I was looking for throughout the duration of the study. When they told me about events that had happened longer ago, these always involved more people, so the data coding could be based on the overlapping part of their stories, and generally some form of documentation was available, too, such as in the shift log book or in work orders, to help in reconstructing the story.

The difference between the studies with regard to both the time elapsed between failure and detection and the performance stage when detection occurred, i.e. the detection occurred later in cases in study 1 than in the cases in the other two studies, corresponds with the finding regarding the relation between preceding failures and moment of detection. After all, study 1 involved more organizational failures, which took longer to be detected.

Finally, the cases in study 2 had lower scores on remaining consequences than the cases in the other two studies, a finding that corresponds with the finding that cases involving human failures only had lower remaining consequences scores.

3.4.3 Limitations of the studies

The main limitation of the three initial studies described in this chapter was the extent to which it was possible to establish causal relationships based on the collected data. Especially in the findings regarding relationships between the variables that can be considered to represent recovery performance and other variables that may have influenced this performance, demonstrating causality is useful. The total number of cases collected in the initial three studies was large enough to allow the use of statistical analyses to identify relationships or patterns in the data, as long as only two variables were considered at a time.

This was already a fortunate situation for a research design using case studies. Case studies often involve too few cases to be able to use any statistical methods at all, or such methods may not be applicable for the data (Yin, 1994). Even though the existence of relationships between various pairs of variables was demonstrated statistically, for many of these relationships, however, given the nature of the dataset, only relatively tentative conclusions could be drawn regarding whether it involved a direct, causal relationship, or if and how other factors also played a role in the relation. To draw such tentative conclusions, I had to rely on methods such as time series analysis and pattern matching, proposed by Yin (1994) as techniques that can be used in case studies to demonstrate causality. In line with these methods, based on the moment in time when something occurred or became known, I combined the identified relations into larger patterns, covering several relationships, complete with tentative conclusions regarding the existence and direction of causality in these relations.

Taking into account the combined effect of several factors, instead of considering them one by one, is not only important for the variables that represent recovery performance, such as remaining outcomes and time-related issues, but also for establishing how tasks are prioritized and distributed in situations where recovery is needed. However, given what was recorded during data collection concerning the priority given to recovery versus other tasks, the distribution of tasks when recovery tasks are added to the existing workload, and factors that were expected to be relevant for decisions made in this regard, it was not possible to analyze the combined effect of all these factors on task prioritization and distribution.

By limiting myself to the variables I used so far for the assessment of performance with regard to recovery and for determining what led to that performance, I have taken a simplified look at what actually goes on. During the three initial studies it became clear that a variety of additional factors could play a role as well. As was indicated earlier, an example of a variable that was not included in the studies, that is, about which no information was recorded, but which could serve as an additional indicator of recovery performance (and of priority given to recovery), is how much time passes between detection and the first recovery step that follows. Other variables that were not included are the quality of performance in separate recovery steps, that is, to what extent was the goal that was set for that step achieved. Failing to achieve a goal set for a separate step will result in the need for additional steps and will therefore at least result in a longer recovery process. Related to this variable is the completeness of the explanation for the deviation that was established by those involved in recovery. The occurrence of explanation steps in itself is no guarantee that these were successful. And the absence of explanation steps does not mean that the causes and extent of the deviation are unknown, they may have been obvious at the moment of detection. Having a complete explanation, no matter how this was achieved, may also influence recovery performance.

3.4.4 Generalizability

To optimize generalizability and thus the external validity of the findings from the initial three studies, as explained in various parts of section 3.2, the principles of replication (Yin, 1994) were used to select organizations and cases to be studied. For reasons of access, and to avoid the possibility that differences in findings between organizations might be the result of the industry in which they operate, only chemical plants were involved in the studies. Care was taken to include as many different recovery processes as possible with regard to all the aspects that were studied. This was the reason to perform not only the second study specifically aimed at adding processes involving self-made errors to the dataset, but also the third study involving processes in a different organization. As a result, it is not very likely that within the domain of the chemical process industry cases can be found that have no similarities at all with any of the cases included in the studies.

Many of the findings are likely to be applicable in other high reliability organizations as well. Their approaches in safety and reliability, after all, are comparable. The recovery process phase model as well as most of the variables that together characterize the recovery processes are flexible enough to accommodate variations between processes, as they had to do so already within the initial three studies. Furthermore, the phase model itself describes processes followed by individuals and groups, and individuals and groups in other organizations may be expected to show similar behavior. The process phase model and the variables used to characterize the recovery processes can serve as tools for organizations to investigate the recovery processes that take place inside their boundaries, as was done in the three initial studies. If the time scales applicable in the development of incidents in other high reliability organizations' processes and the possible means for recovery are very different from those in the chemical industry (which, in fact, already involves already a wide range of time scales), however, the occurrence frequencies of the different types of recovery processes may be different.

In addition, some of the findings could be equally relevant for non-high reliability organizations, for example in health care (e.g. Kohn, Corrigan & Donaldson, 2000), too. Although the safety and reliability performance of those organizations may not be at the same high level, promoting recovery possibilities still provides an additional way to improve their safety and reliability. In fact, this may be an important step to eventually becoming a high reliability organization. A study in two different medical settings (Van

Vuuren & Kanse, in press) has already demonstrated that the process phase model and the variables characterizing the recovery processes were equally applicable in that domain.

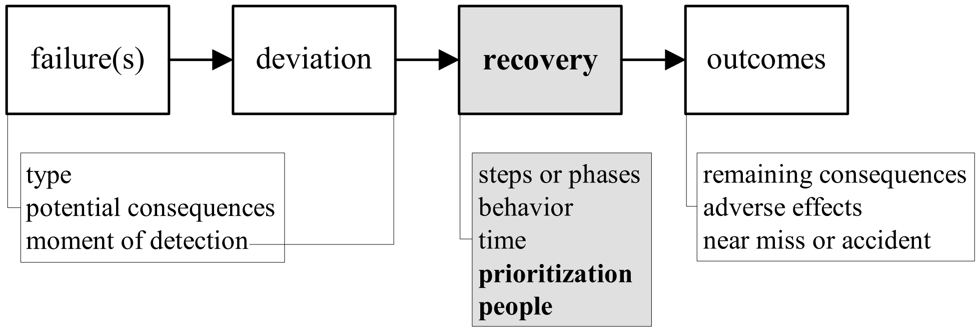
3.4.5 Future research directions

One of the most obvious areas needing additional research was related to research questions 4 and 5, concerning task prioritization and distribution in situations where recovery is needed. The operationalization at the time of data collection in the initial three studies of the priority given to recovery tasks and the factors expected to influence task prioritization and distribution turned out to be unsuitable for a complete analysis, especially for an *integrated* look at both the prioritization and the distribution of tasks together, given a certain *combination* of influences. After all, most often the way tasks are prioritized and distributed is not just the result of the influence of merely one factor, which is how things were treated in this chapter, but of a combination instead. As a result, the answers to research questions 4 and 5 were not as complete as they could have been. Therefore, I chose to conduct two additional empirical studies, specifically aimed at a more in-depth, systematical analysis of the combined effect of the various influences on both task prioritization and task distribution in situations where recovery is needed. These studies and the resulting findings are described and discussed in chapter 4.

The findings from the research described so far also pointed towards other, more general areas worth exploring in more detail. Among these areas is the multitude of factors, besides those influencing task prioritization and distribution, that influence recovery processes, separate phases, transitions between phases, time involved, and process outcomes –from my data it was obvious that these factors are not limited to those I have specifically analyzed so far. To establish which of such factors are most influential, research on how to measure the success of a recovery process other than by simply focusing on remaining consequences could be beneficial. Another area worth exploring concerns the use of more sophisticated statistical techniques preferably on a larger dataset, to establish which interrelations exist between multiple recovery process-related factors, so that in the analysis of a relationship between two factors other factors are no longer disregarded. Also, the current findings should be extended to include other types of organizations. However, within the scope of this thesis there was no room to follow up on all the possible areas for additional research. After the discussion of the two follow-up studies in the next chapter, in chapter 5 I will give a detailed overview of what I consider the most promising areas.

4 Follow-up studies

4.1 Purpose and outline



As was pointed out already at the end of the previous chapter, the findings from the initial three studies that were performed to answer the research questions posed in this thesis indicated a need for further research, among other things, with regard to the fourth and fifth specific research questions. Recall from chapter 2, subsections 2.2.4 and 2.2.5, that these questions concerned task prioritization and distribution of tasks to others during situations in a working environment where a failure or a combination of several different or similar types of failures has led to a deviation that needs to be recovered:

> *Based on which factors do people prioritize tasks when a deviation needing recovery presents itself at work?*
> *If others than those already involved become involved in a recovery process at work:*
> *a) what are the organizational roles of these people,*
> *b) when do they become involved,*
> *c) for what reasons do they become involved, and*
> *d) what is the effect of their involvement?*
> *Furthermore, if those involved in detection do not involve any others in the rest of the recovery process at all, for what reasons does this happen?*

In this chapter, two follow-up studies are presented, that were performed to gain a more complete and detailed understanding of the roles played by several factors in prioritization and distribution of tasks in situations where recovery tasks are added to people's already existing workload. The chapter starts with a theoretical section, 4.2. In this section, first the insights available from existing literature regarding the influence of situation-related factors on task prioritization and distribution, that were presented already in chapter 2, are repeated and combined with the insights gained in the initial studies regarding these factors, that were presented in chapter 3. These insights are summarized in a conceptual model, and expectations are formulated regarding the effects of these situation-related factors on task prioritization, task distribution, and the extent to which this distribution differs from what the organization prescribes. After that, insights from additional literature are introduced, concerning the effects of person-related (as opposed to situation-related) factors on task prioritization, task distribution, and the extent to which this distribution differs from what the organization prescribes. The conceptual model is updated accordingly. Then, section 4.3 follows, in which the hypotheses and additional research questions that were formulated based on all of these insights are presented. Section 4.4 addresses the research design for the follow-up studies that were performed to test these hypotheses and answer these additional research questions. Both issues related to data collection and to data analysis are discussed in this section. After that, in the first five subsections of section 4.5, the findings from the studies are presented and discussed. And finally, in the last three subsections of section 4.5, the conclusions are summarized, the limitations of the studies are discussed, and some possible directions for future research are given.

4.2 Theory and conceptual model

4.2.1 Situation-related factors

The idea that people may (re-)assess task prioritities when confronted with a situation that adds recovery tasks to the already existing workload, and that they may also (re-) distribute tasks to deal with this situation, is based on findings from, for example, Waller (1999, 1997), Maule (1997), and Kecklund and Svenson (1997). Waller (1999) found task prioritization to be one of the ways in which people respond to non-routine events, and referred to the possibility of re-allocating or distributing (sub)tasks (Waller, 1997) to other group members in multitask situations. Maule (1997) and Kecklund and Svenson (1997) also mentioned task delegation as a way to adapt to time pressure or to cope with demanding work situations. Since task prioritization and task distribution are both adaptive responses of those who are confronted with a situation in need of recovery, it may be best to study those responses in combination. Furthermore, the factors that influence one of these adaptive responses may also influence the other. In the literature review in subsections 2.2.4 and 2.2.5 and in the corresponding findings from the initial studies in subsections 3.3.4 and 3.3.5, a number of situation-related factors were presented that were found or at least expected to influence how tasks are prioritized and/or distributed across different people when a situation needing recovery presents itself. These factors were:

- What written or unwritten protocols, procedures or work instructions prescribe or suggest with regard to who should perform the necessary recovery tasks.

In subsection 3.3.5, table 14 and below, this factor was shown to be an important reason for involving others. This factor may also influence task prioritization, for example, it may take some time before the people with the required authorization or qualifications arrive on the scene.

- The availability of the required knowledge and skills or expertise.
  Waller (1997) mentioned this factor as a reason for re-allocating tasks to others. Heckhausen and Kuhl (1985) argued that having the necessary means plays a role in the transformation of intentions into actions. Knowledge, skills and expertise can be seen as means needed to perform certain tasks or actions. Furthermore, in subsection 3.3.5, table 14 and below, the availability of the required knowledge and expertise was shown to be an important reason for involving others.
- The availability of the necessary tools and materials or equipment.
  As was stated before, Heckhausen and Kuhl (1985) mentioned having the necessary means as a factor in the transformation of intentions into actions. Tools and materials or equipment can also be considered as means required to perform certain actions. In addition, in subsection 3.3.5, table 14 and below, this factor was shown to be an important reason for involving others.
- The urgency of performing recovery steps, defined as the time available until the earliest possible occurrence of negative consequences if no recovery steps would be taken.
  Geen (1995) argued that more urgent actions get precedence over other actions. Time management research by, for example, Covey et al. (1994) and Koch and Kleinmann (2002), also demonstrated that urgency plays an important role in task prioritization. In addition, Heckhausen and Kuhl (1985) mentioned urgency as a factor in the transformation of intentions into actions. Furthermore, Hollnagel (2002) also underlined the importance of the available time for actions. This factor may also affect task distribution, as people who detect a deviation for which recovery is urgent may want to speed up the process by getting started on it immediately, personally. Or if they cannot stop their ongoing tasks, they may want others to get started on it immediately, or to take over their ongoing tasks. In either case more people will be needed.
- The importance of performing recovery steps, defined by the types and severity of the potential negative consequences (i.e. not the same as the urgency); the larger the number of and the more severe the potential consequences are, the more important is recovery.
  Time management research (Covey et al.,1994; Koch & Kleinmann, 2002) has shown that the importance of a task influences the priority given to that task. Also according to Heckhausen and Kuhl (1985), importance is a factor in the transformation of intentions to actions. This factor may also influence task distribution, in that people

who detect a deviation for which recovery is important may feel that the potential consequences are so daunting that they want to get started on the recovery process in person, immediately. Or, similar to what was argued for urgent recovery tasks, if they cannot stop their ongoing tasks, they may want others to get started on it immediately, or to take over their ongoing tasks. Again, in either case more people will be needed.

- The amount of time needed for recovery steps.
  This factor was mentioned by Heckhausen and Kuhl (1985) as another factor in the transformation of intentions into actions. Hollnagel (2002) pointed out that problems arise when this amount of time is more than the available time. Furthermore, in subsection 3.3.5, table 14 and below, not having the time that was needed for the recovery tasks was shown to be one of the reasons for involving others.
- The presence of other tasks requiring immediate attention.
  Similar to what was argued based on findings by Geen (1995), Covey et al. (1994), Koch and Kleinmann (2002), Claessens et al. (2003), and Heckhausen and Kuhl (1985), with regard to the urgency and importance of recovery tasks, also for the tasks on which people are already working when recovery tasks are added to the workload, priorities will be (re-)assessed based on their urgency and importance. These other tasks may also require immediate attention, and may actually get a higher priority than the newly added recovery tasks. In the reasons found in the initial studies for involving others, presented in subsection 3.3.5, table 14 and below, the presence of other tasks requiring immediate attention may have played a role both in not having the time that was needed for the recovery tasks, and in not having sufficient resources. As there are only so many tasks a person can handle at the same time, if both the tasks on which one was already working and the newly added recovery tasks require immediate attention, the assistance of others may be enlisted.
- The availability of others who can assist.
  This factor is related to what Heckhausen and Kuhl (1985) argued with regard to having the means needed to perform the required tasks, that this factor influences the transformation of intentions into actions. These means, after all, can take the form of knowledge and skills, tools and equipment, or people, that is, human resources or manpower. In subsection 3.3.5, table 14 and below, human resources were also shown to be a reason for involving others in recovery tasks. If there are no others available who can assist, it is not possible to re-allocate any tasks to them that require immediate attention, neither recovery tasks nor already ongoing tasks.

As can be seen from this overview, these situation-related factors, based on which people decide what priority tasks get and who should be involved, encompass both (written or unwritten) rule- or procedure-related reasons, recovery tasks-related reasons, and resource-related reasons. Summarizing all of the above, figure 6 contains a conceptual model that shows which of the factors belong in which of these three categories of reasons, and depicts the influences of the factors on both task prioritization and task distribution in situations where recovery is needed.

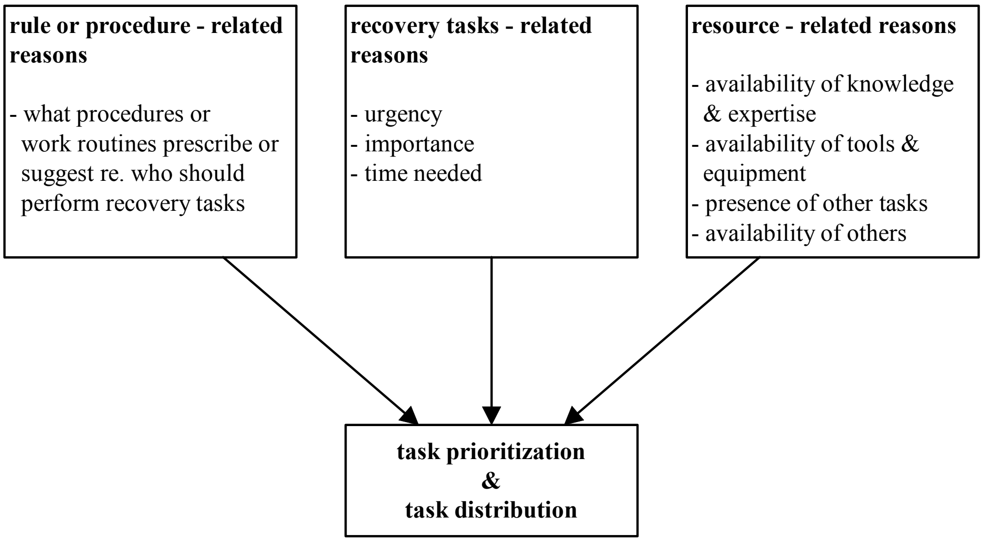


Figure 6 Situation-related factors that influence task prioritization and distribution

4.2.2 Expected effects from situation-related factors on task prioritization

Regarding task *prioritization* in situations where recovery from a deviation is needed, two types of tasks can be distinguished: the *recovery tasks* that are newly added to the to-do-list when a deviation is detected, and the *tasks on which people who detect a deviation are already working* at that moment. In the research described here, the focus will be on the priority given to recovery tasks. The degree to which the already ongoing tasks also require immediate attention (i.e. the priority these deserve) is regarded as one of the given, situation-related factors that can influence the priority given to recovery tasks.

From all of the factors summarized in figure 6, urgency and importance are expected to be most influential on the priority given to recovery tasks. Urgent and important recovery steps are expected to be associated with high priorities given to these steps (Heckhausen & Kuhl, 1985). In those circumstances recovery steps are expected to start soon after a deviation is detected. Urgency may even be more important in prioritizing than importance (e.g. Koch & Kleinmann, 2002; Claessens et al., 2003). The rest of the factors included in the model shown in figure 6 may either extend or shorten the time period between detection and the first recovery step that follows, depending on their values. If the written or unwritten (explicit or implicit) procedures prescribe or suggest that others should be involved in the recovery process, especially when a complete handover is suggested, this may cause some delay. If people do not have the required knowledge and expertise, this may cause some delay, as they then need to find people who are able to perform the required tasks, and these people may not immediately be available. Similarly, if the necessary tools and equipment are not available, this may cause some delay, as these need to be collected from elsewhere. If relatively much time is needed to complete the recovery steps, this may also cause some delay, as most people have a tendency to perform actions that can be done quickly in between other tasks on which they were already working, but tend to postpone the lengthier tasks. This corresponds with predictions from Koch and Kleinmann (2002) regarding the effect time discounting in task prioritization. They argue that people tend to prefer tasks of which the goal can be achieved within a short time. If people who detect a deviation are at that moment already busy with other urgent and important tasks, this may cause some delay, as these ongoing tasks cannot easily be put on hold. They compete for attention with the recovery steps that from that moment onwards also require attention. And finally, if other people who can assist are not available, this may cause some delay, as further recovery steps have to wait until their arrival. This may especially be so in cases where the recovery tasks cannot be performed by those who detect a deviation alone, or in cases where those who detect a deviation are already very busy with other urgent and important tasks.

4.2.3 Expected effects from situation-related factors on task distribution

Also with regard to the *distribution* of tasks in situations where recovery is needed, two types of tasks can be distinguished. Again, these tasks are the *recovery tasks* that are newly added to the to-do-list when a deviation is detected, and the *tasks on which people who detect a deviation are already working* at that moment. In the research described here, the focus will be on the distribution of recovery tasks. The degree to which the already ongoing tasks normally require immediate attention is regarded as one of the situation-related factors. No specific attention is given to how these tasks are distributed. However, the choice people make regarding their further role in the task on which they were already working when confronted with a deviation in need of recovery, can to a reasonable extent be inferred from the share of the recovery tasks they chose to take on personally, the moment the recovery tasks commence (i.e. priority given to recovery), and the degree to which these already ongoing tasks require immediate attention. For example, if people decide to perform the recovery tasks by themselves and start immediately, and the ongoing tasks also require immediate attention, they are likely to hand over these ongoing tasks to others, or at least to some extent, until they have finished the most urgent recovery steps.

How the recovery steps required after a deviation is detected are distributed, or by whom these recovery steps are performed, or, more simply, the extent to which others are involved in all of the necessary recovery steps, may depend on the same factors as those expected to influence the moment when the first recovery steps after the detection of a deviation are started. The involvement of others is in principle dictated by written or unwritten (explicit or implicit) procedures or work routines – if these prescribe or suggest that cooperation with others is needed, or even a complete handover to others, then the extent to which those who detect a deviation are involved in the rest of the recovery steps is expected to be lower or in the latter case possibly even reduced to zero, than in cases where no such cooperation or handover is suggested. If the people who detect a deviation do not possess the knowledge and expertise required to perform the necessary recovery steps, the extent to which others are involved in the recovery steps may be larger, that is, others whom they expect to have such knowledge and expertise, to give the recovery process a better chance to be successful. If the necessary tools and equipment are not readily available to those who detect a deviation, this may also decrease the extent to which these persons are involved and increase the extent to which others are involved in the required recovery steps. This may happen for different reasons: the others whose assistance is recruited may be the ones who have the necessary tools and equipment, or they may be the ones who are on duty at the moment when the required tools and equipment become available. The latter situation is possible with shift work in a continuous production environment. Furthermore, if the recovery steps are urgent, people who detect a deviation may be involved to a larger extent in its subsequent recovery than in less urgent cases, and the same may apply for important recovery steps. After all, these people are already present and know about the deviation, which will save time otherwise needed to inform people before any steps can be started. If much time is required to perform the necessary recovery steps, this may lead to more involvement of others and less of the people who detected the deviation. While people may be able to quickly perform a small task in between other tasks on which they were already working, it may be problematic to accommodate lengthier tasks, so for these the assistance of others may be sought. If the other tasks on which the people who detect a deviation are working are urgent and important, this may lead to more involvement of others and less of themselves in the necessary recovery steps. They may simply be too busy to deal with additional tasks. Another option they have is to recruit others to work on the tasks they were engaged in at the moment of detection. Finally, if people who can assist are not available, the people who detect a deviation may take a larger share of the recovery steps on themselves than in cases where assistance is available, to avoid potentially harmful delays.

4.2.4 Expected effects from situation-related factors on the extent to which the distribution of recovery tasks differs from what the organization prescribes

Derived from the extent to which people decide to involve others in the required recovery steps is whether people, in this decision, deviate from the normal procedure or work routine (documented or undocumented) with regard to who should be involved in the recovery steps. Note that the words 'deviate' and 'deviation' are now also used to indicate situations where the normal routines or procedures for recovery are not followed exactly. This use of these words should not be confused with the meaning of the word deviation that was introduced in chapters 1 and 2, to indicate a problem situation that is considered unacceptable and that needs to be recovered. To distinguish between these two meanings of the words 'deviate' and 'deviation', if they are meant to describe a deviation from the normal recovery procedures, this is always indicated in the text. A final remark about procedure deviations is that in organizations, over time, some deviations from formal (written) procedures may have become generally accepted practice, and thus informal procedures or work routines. In the studies described in this chapter, if both formal and informal procedures existed with regard to who should be involved in certain recovery tasks, the formal procedures were used as point of reference. As a result, if in such cases, that is, in such scenarios, the collected data would reveal that people deviated from normal procedures regarding who to involve in the recovery steps following detection, this could either be indicative of a general, collective tendency or merely an individual one. In those scenarios, a closer inspection of all the responses given by different participants to the same scenario could reveal which of the two tendencies played a role. Such an analysis at the level of separate scenarios, however, was outside the scope of this chapter.

A deviation from the normal procedure with regard to who should be involved in recovery is generally not a problem situation that needs to be recovered. In some cases, however, such a procedure deviation might be problematic, for example if people end up performing tasks while lacking the appropriate knowledge, or without the proper tools. A better insight in the factors that may lead people to deviate from procedures with regard to how recovery tasks should be distributed can show if such unwanted cases are likely. Furthermore, such insight can also show under which circumstances specific procedures regarding who should perform certain recovery tasks may not be practical or not even feasible, and can thus help to improve existing procedures.

If a deviation from the normal procedures regarding who should be involved in the recovery steps after detection occurs, this can be in two directions: either people take on more than their suggested share of the recovery steps, or they take on less. The same factors as those that are expected to influence recovery task prioritization and distribution may also influence the extent and direction of the deviation from procedures or work routines regarding who should perform the recovery steps. What the normal procedure suggests regarding who should be involved in recovery may influence the extent and direction of the procedure deviation. If the procedure suggests a person can perform a recovery task alone, the only possible procedure deviation is to take on less than the suggested share of the work, and thus to involve others to some extent. If the procedure suggests that a task has to be handed over completely to others, the only possible procedure deviation is to take on more than the suggested share of the work. The greater the suggested role of others in recovery, the greater may also be the procedure deviation by those who detect a deviation in the direction of taking on more than the suggested share, especially in processes where the operators are responsible for safe operations. Operators are not expected to be the kind of people who 'sit back and watch' things go wrong if they can help it.

Having the required knowledge and expertise may also be an important precondition in the decision to take on a larger share of recovery steps than suggested. Without that knowledge, even if people are willing to do something other than to recruit the assistance of others, they may not be able to. The same applies to having the necessary tools and equipment available. With these present, people may be more inclined to take on a larger share of the recovery steps than suggested by procedures or work routines. The more urgent and more important recovery steps are, the more inclined those who detect a deviation may be to take on a larger share of the required recovery steps than they are supposed to. They may feel that that is the only way to keep the plant and process safe. For less urgent or less important recovery steps, they may feel far less need to deviate from the procedures in the direction of doing more themselves. There may then even be circumstances under which they would deviate from the procedure in the opposite direction, i.e. involving others to a larger extent. If the time required for the recovery steps is long, people may be more inclined to deviate from the suggested routine in the direction of letting others play a larger role. After all, they may not even be on duty any more to see the end of all the required recovery steps, or their attention may be required elsewhere after a while. If the people who detect a deviation are busy at that moment with other urgent and important tasks, they may also be inclined to deviate from the suggested routine in the direction of letting others play a larger role, as this may then be the most efficient way to deal with multiple tasks. And finally, if no others who can assist are available, people who detect a deviation may take on a larger share of the required recovery steps than suggested by the procedures or work routines, simply because something needs to be done to avoid negative consequences and they may be the only one available to do so.

4.2.5 Person-related factors and their effects on task prioritization and distribution

In addition to *situation-related factors*, *person-related factors* may also influence recovery task prioritization and distribution, and the deviation from what the normal, explicit or implicit procedures prescribe regarding who should perform the required recovery tasks.

In work- and organizational psychology, a person's *self-efficacy* is often used to explain interpersonal differences with regard to work-related issues such as performance and job satisfaction, as was pointed out by, for example, Schyns and Von Collani (2002). Self-efficacy is defined as a person's conviction that he or she can successfully execute a given behavior required to produce certain outcomes (cf. Bandura, 1977). According to Bandura (1997), the outcomes that people anticipate depend largely on their judgments of how well they will be able to perform in given situations, that is, their self-efficacy. This means that people who believe that they do not have the ability to perform a certain task well expect to be unsuccessful in reaching the outcomes associated with that task. Therefore, they may be less likely to take on that task in the first place. The most common method to measure a person's self-efficacy is via a questionnaire. To this end, over the years, several scales have been developed and validated (e.g. Bandura, 1977; Sherer, Maddux, Mercandante, Prentice-Dunn, Jacobs & Rogers, 1982; Shelton, 1990; Eden & Kinnar, 1991; Chen, Gully & Eden, 2001), including a scale specifically aimed at measuring work-related or occupational self-efficacy (Schyns & Von Collani, 2002). Schyns and Von Collani (2002) argued that the specificity of a measuring instrument for a predictor variable should match the specificity of the outcome variable that is being considered. They developed the occupational self-efficacy scale for specific use in work environments, thereby not restricting its applicability to certain types of tasks only, but just occupational settings in general. Their occupational self-efficacy scale has been validated in at least three studies (Schyns & Von Collani, 2002). This scale may also be helpful in explaining interpersonal differences in the way persons decide to prioritize and distribute tasks in situations at work where recovery is needed, and in their willingness to deviate from what the normal, explicit or implicit procedures prescribe with regard to who should be involved.

With regard to how soon recovery steps will start after a deviation is detected, i.e. the priority given to recovery, it is hard to say how this may be correlated with a person's occupational self-efficacy. Persons with high scores on occupational self-efficacy may start later with the first recovery step after detecting a deviation than persons with lower scores, believing that they can do what is required quickly and without error, that they don't need any additional time to allow for several attempts, while people with low scores on occupational self-efficacy may want to build in some extra time. But the opposite may be true as well, that persons with high scores on occupational self-efficacy may start earlier than people with low scores. This may be so because people with high occupational self-efficacy scores prefer dealing with challenges rather than their normal tasks. Or because people with low self-efficacy scores think they will not be able to correct the problem and put off getting started on it, thinking that there may be no use in getting started, or hoping someone else will deal with it if they wait a while.

With regard to the extent to which others will be involved in the recovery steps, persons with higher occupational self-efficacy scores may be less inclined to do this than persons with lower self-efficacy scores, as the former may feel they will be able to handle things without too much assistance, compared to the latter who are not so sure they can, and may prefer to involve others.

As for the deviation from what the normal, explicit or implicit procedures prescribe or suggest with regard to who should be involved in the recovery steps after detection, persons with higher occupational self-efficacy scores may be more inclined to take on more than their share of recovery tasks than persons with lower self-efficacy scores. Even though they are officially not required or supposed to do so, the former may feel that they can handle these tasks, while the latter may actually even hand over their official share in the recovery tasks to others, not being sure they can handle this. There may be limits to the described tendencies, however. For some recovery tasks, even people with high occupational self-efficacy scores may think that they are best left to others.

Another factor that may explain interpersonal differences in the way people decide to prioritize and distribute tasks in situations where recovery is needed and in their willingness to deviate from what the normal, explicit or implicit procedures prescribe with regard to who should be involved, is how many years of *experience* persons have in their current line of work. How the number of years of relevant work experience correlates with the priority given to recovery, the extent to which others are involved in the recovery tasks, and the willingness to deviate from normal procedures regarding who should be involved in recovery, is difficult to predict. More experience may have made people better prepared for whatever situations they are confronted with. As was explained already in chapter 2, subsection 2.2.2, with experience, for some tasks the information processing behavior level necessary to perform the task may shift from knowledge-based, to rule-based or even skill-based behavior (cf. Rasmussen, 1976, 1986), that is, a shift from non-routine towards routine, or even automatical behavior. More experience can either lead to a more relaxed attitude, starting the necessary recovery steps later after detection, taking on more tasks, possibly even tasks in which others should be involved, or to a better awareness of risks involved, triggering an early start of the recovery steps, restricting the amount of tasks one is involved in, not overstepping the boundaries of one’s responsibilities and abilities.

The difficulties as described above to predict which correlations may exist between the person-related factors, that is occupational self-efficacy and years of relevant work experience, on the one hand, and recovery task prioritization, distribution, and willingness to deviate from the procedures regarding who should be involved, on the other hand, suggested an exploration of the possible interactions between the person-related factors and situation-related factors. That way, it would be possible to establish whether there are effects that only exist for certain combinations of person- and situation-related factors. As an example, occupational self-efficacy may only affect the extent to which people decide to involve others in recovery, if such others are actually available.

In figure 7, the conceptual model of figure 6 is updated to not only include the situation-related factors, but also the person-related factors that influence task prioritization and task distribution in situations where recovery is needed.

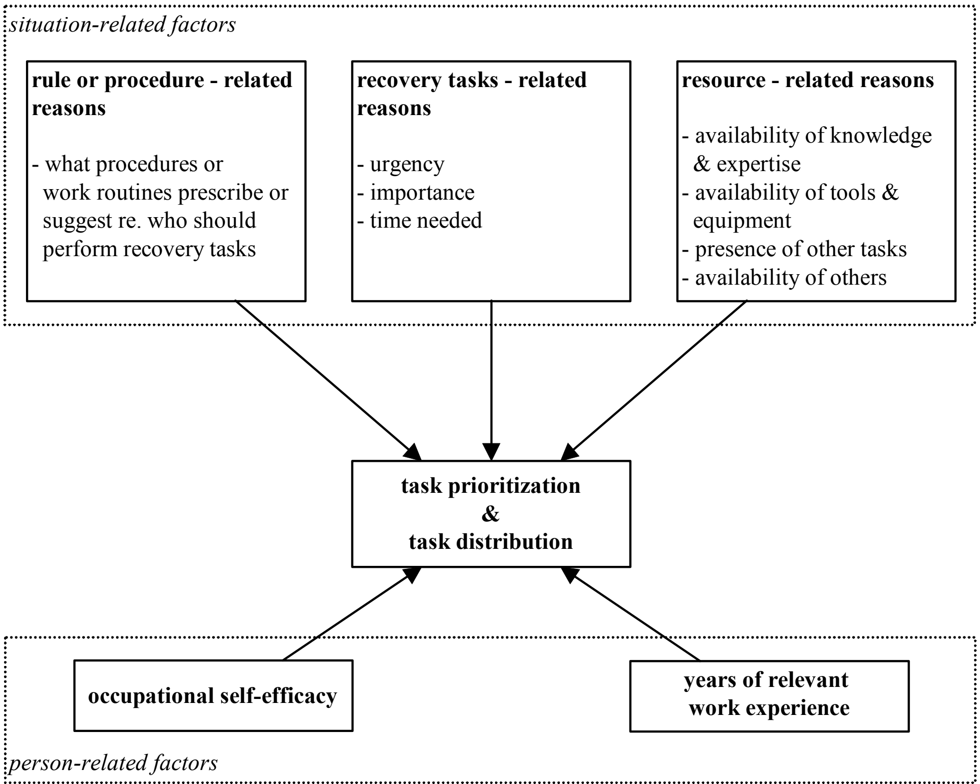


Figure 7 Situation and person-related factors that influence task prioritization and distribution

4.3 Hypotheses and research questions

Everything that was discussed so far in this chapter can be summarized in the following hypotheses and research questions:

Hypothesis 1:
The following situation characteristics each increase the amount of time that passes after the detection of a deviation before the first recovery step is started:

a) the normal, explicit or implicit procedures prescribe or suggest that others than those detecting the deviation have to be involved in the recovery steps (the larger the suggested role of these others, the larger the delay),
b) the people who detect the deviation do not have the required knowledge and expertise to deal with it,
c) the people who detect the deviation do not have the necessary tools and equipment,
d) the recovery steps are not so urgent,
e) the recovery steps are not so important,
f) much time is needed to perform all of the recovery steps,
g) the people who detect the deviation are busy with other urgent and important tasks,
h) others who can assist are not available,

whereas opposite characteristics each decrease this amount of time.

Hypothesis 2:
The following situation characteristics each increase the extent to which people who detect a deviation decide to involve others in the recovery steps that follow:

a) the normal, explicit or implicit procedures prescribe or suggest that others than those detecting the deviation have to be involved in the recovery steps (the larger the suggested role of these others, the larger their actual involvement),
b) the people who detect the deviation do not have the required knowledge and expertise to deal with it,
c) the people who detect the deviation do not have the necessary tools and equipment,
d) the recovery steps are not so urgent,
e) the recovery steps are not so important,
f) much time is needed to perform all of the recovery steps,
g) the people who detect the deviation are busy with other urgent and important tasks,

h) others who can assist are available,
whereas opposite characteristics each decrease this extent.

Hypothesis 3:
Deviations are possible from what the normal, explicit or implicit procedures prescribe or suggest with regard to who should be involved in the recovery steps that follow detection. The following situation characteristics each shift a possible procedure deviation in the direction of those involved in detection themselves taking on a larger share of the recovery steps than prescribed or suggested:

a) the normal, explicit or implicit procedures prescribe or suggest that others than those detecting the deviation have to be involved in the recovery steps (the larger the suggested role of these others, the larger the shift in the direction of those involved in detection themselves taking on a larger than suggested share of the recovery steps),
b) the people who detect the deviation have the required knowledge and expertise to deal with it,
c) the people who detect the deviation have the necessary tools and equipment,
d) the recovery steps are urgent,
e) the recovery steps are important,
f) not much time is needed to perform all of the recovery steps,
g) the people who detect the deviation are not busy with other urgent and important tasks,
h) others who can assist are not available,

whereas opposite characteristics each shift the possible procedure deviation in the direction of others than those involved in detection taking on a larger share of the recovery steps than prescribed or suggested.

Research question A:
What are the effects of a person's occupational self-efficacy on

1. *the moment when the first recovery step is started after the detection of a deviation,*
2. *the extent to which people who detect a deviation decide to involve others in the recovery steps that follow,*
3. *the deviation from what the normal, explicit or implicit procedures prescribe or suggest with regard to who should be involved in the recovery steps that follow detection,*
4. *and which interaction effects exist of the person-related factor occupational self-efficacy with situation-related factors, for each of these three variables?*

Research question B:
What are the effects of the number of years of a person's relevant work experience on

1. *the moment when the first recovery step is started after the detection of a deviation,*
2. *the extent to which people who detect a deviation decide to involve others in the recovery steps that follow,*
3. *the deviation from what the normal, explicit or implicit procedures prescribe or suggest with regard to who should be involved in the recovery steps that follow detection,*
4. *and which interaction effects exist of the person-related factor years of relevant work experience with situation-related factors, for each of these three variables?*

4.4 Research design

4.4.1 Selecting organizations

The reasons given in subsection 3.2.2 for choosing the chemical process industry as the domain where recovery processes would be studied were also applicable for the selection of organizations to participate in the follow-up studies that were needed to test the hypotheses and answer the research questions presented above. In summary, these reasons involved the presence of and familiarity with recovery processes within organizations operating in this domain. What is more, staying in the same domain would minimize the chance that potential differences in findings would be the result of the change to a different domain.

Following the replication principles (cf. Yin, 1994) outlined in subsection 3.2.1, I approached different chemical plants to participate in the follow-up studies than those who had already participated in the initial studies. Including different organizations would help to ensure external validity and generalizability of the findings. In the follow-up studies, again two chemical plants participated, one in study 4, and one in study 5. Both were based in the Netherlands, both were local branches of a (different) multinational enterprise. They both had similar production processes and made the same types of product, that is, the plants consisted of catalytic cracking units, where naphtha was 'cracked' into different smaller hydrocarbon chains for use in other production processes. As a result of their similarities, the same research design could be used in both organizations. This not only speeded up the research design process itself, but it also allowed combining the two resulting datasets into a larger one, which was beneficial for statistical analyses and for demonstrating generalizability of the findings.

4.4.2 Vignette study approach

One of the disadvantages of the data collection approach used in the initial three studies was that for all of the recovery processes on which data was collected, data were only available regarding the response of those who were actually involved in the recovery steps to the situation they had to deal with. As every event in the dataset was unique, it was unknown whether other people might have responded differently to the same situation. In addition, not all the possible combinations of factors that were expected to influence task prioritization and distribution were represented in the dataset. To overcome these problems I decided to use a different research method in the follow-up studies, a design whereby I could control the deviations to which study participants had to respond. Several methods could be used to this end, including the use of simulators in which operators would be confronted with a simulated deviation. While this could have guaranteed realistic representations of possible deviations involving control room operator tasks, it would have been less suited to represent the tasks of field operators who often work in the plant, away from the process control computer interfaces. Also, developing simulated deviations is not only a very time-consuming but also very costly process. Therefore, I chose a more simplified approach to simulate deviations, by describing them in text-form, thus creating written scenarios. These scenarios were presented to operators who were asked to respond to a number of questions for each scenario. The approach I used is known as a vignette study or a factorial survey (e.g. Rossi & Anderson, 1982), in which the scenarios described above are referred to as 'vignettes' or 'factorial objects'. Vignette studies or factorial surveys have often been used for decision and judgment research, predominantly in social sciences (e.g. Hox, Kreft & Hermkens, 1991; Rossi & Nock, 1982). They combine characteristics of research involving factorial experiments and research using surveys. Vignette studies are different from full factorial experiments in that there is no need to have all participants respond to all possible combinations of the different levels of the factors included in the experiment, they merely need to respond to a sample of all the vignettes. Therefore, a much larger number of factors and levels within factors can be included. Vignette studies are also different from surveys in that they enable a researcher to distinguish separate effects of factors that can not normally be separated in surveys as a result of being closely related to each other. Vignette studies allow the analysis of both the effects of the factors built into the vignettes and of the person-related factors of participants in the study (Rossi & Anderson, 1982; Hox et al., 1991).

4.4.3 Vignette development

The factors I included in the vignettes were the eight factors listed at the beginning of section 4.2 and repeated later on in figure 6, in the three boxes at the top of figure 7, and in hypotheses 1, 2 and 3. They are listed again below, including an indication of the different levels I distinguished for each of them:

a) what written or unwritten protocols, procedures or work instructions prescribe or suggest with regard to who should perform the necessary recovery tasks (0=entirely by oneself, 1=share with others, 2=complete handover to others),
b) the availability of the required knowledge and skills or expertise (0=knowledge and expertise not available, 1=knowledge and expertise available),
c) the availability of the necessary tools and materials or equipment (0=tools and equipment not available, 1=tools and equipment available),
d) the urgency of recovery steps, i.e. the time available until the earliest possible occurrence of negative consequences (0=not so urgent, 1=very urgent),
e) the importance of recovery steps, i.e. the types and severity of the potential negative consequences (0=not so important, 1=very important)
f) the time needed for recovery steps (0=not much time, 1=much time),
g) the presence of other tasks requiring immediate attention (0=other tasks not urgent and important, 1=other tasks urgent and important),
h) the availability of others who can assist (0=no others who can assist available, 1=others who can assist available).

To save on the time a participant would need to 'take in' a certain scenario, I based the vignettes on the first six factors *a)*, *b)*, *c)*, *d)*, *e)*, and *f)*. I used factors g and h to describe the circumstances under which the deviation presented in the vignette could occur. I asked participants to respond to the same vignette, composed of certain levels of factors *a)* to *f)*, four times, that is under four different circumstances, made up by the four different possible combinations of levels of factors *g)* and *h)*. This way, one vignette could be used to collect responses for four different scenarios – scenarios with the same basis of certain levels of factors *a)* to *f)*, but differences with regard to factors *g)* and *h)*.

Theoretically, with this approach still 3*2*2*2*2*2=96 different vignettes were possible – and for each of them four different circumstances. However, not all combinations of levels of the different factors translated into plausible vignettes. For example, situations where the normal procedure suggests that a person who detects a deviation can either perform all necessary recovery steps alone or share them with others, in combination with the required knowledge and expertise not being available, were not very credible or realistic vignettes. Elimination of these improbable vignettes reduced the number to 64 theoretically possible vignettes, with again for each of them four different circumstances.

At the top of each vignette, participants were asked to imagine themselves in the situation that was subsequently described in abstract terms, using bullet points to list which levels of each of the factors *a)* to *f)* were applicable. To ensure that participants would be able to relate to the deviations described in the vignettes, at the bottom of each vignette I included a concrete example of the described situation, that was realistic for their work environment. In the text after the example, the participants were instructed to base their answers on the general situation characteristics provided, not merely on the example alone. I developed the vignettes, specifically the examples, with the assistance of a number of staff members from the plant that participated in study 4, the first follow-up study. All of these persons had ample experience in process operations at this plant. At the time of my study they were employed as training coordinator or safety staff, functions through which they had gained detailed knowledge of all possible deviations at the plant. Altogether we managed to come up with examples for 39 different vignettes. The rest of the possible vignettes were either not considered to be very realistic within the plant, or simply no corresponding examples could be thought of.

Adapting the vignettes so they could be used at the plant that participated in the second follow-up study, study 5, involved only a few minor changes in wording, not in content. Members of this plant's operations- and technological staff[3] reviewed the vignettes for applicability in their plant and made the appropriate changes in the text of the examples. For example, the operators referred to as panel operators in one plant were called control room operators in the other plant, even though their jobs were the same, and occasionally different names were used to refer to the same equipment. Two of the 39 vignettes had to be left out in study 5 as they involved situations that were impossible in the plant that participated in study 5, and no alternative vignettes were added, leaving a total of 37 different vignettes for this study.

In the vignettes, after the example the part that had to be filled in by the participants started. In this part, the same two questions, corresponding with the first two dependent variables in the studies, were asked four different times. Each of these four times the participant was

---

[3] I owe many thanks to Drs. Lesley Smits from the Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, who worked as a temporary member of the technological staff at this plant on a design project, for her assistance in this study.

asked to imagine the described situation occurring under a different set of circumstances. First when he or she would already be busy working on other urgent and important tasks and others who could assist would be available, then when he or she would already be busy working on other urgent and important tasks and *no* others who could assist would be available, then when the tasks on which he or she would be working would *not* be urgent and important and others who could assist would be available, and finally when the tasks on which he or she would be working would *not* be urgent and important and *no* others who could assist would be available. As was explained before, every set of circumstances corresponded with a separate scenario, even though its basis was still the same as for the other three sets of circumstances. The two questions concerned:

1. the moment when the first recovery step is started after the detection of a deviation (three point scale where 1=immediately, 2=later during the same shift, and 3=during a later shift)
2. the extent to which people who detect a deviation would involve others in the recovery steps that follow
   (five point scale where 1=I would do it by myself, no others would be involved, and 5=further recovery would be done completely by others),

The deviation from what the normal procedures or routines prescribed or suggested with regard to who should be involved in the recovery steps that followed detection, the third dependent variable, could be calculated for each scenario, based on the answers given to question 2 as shown above, and the level of factor *a)* in the corresponding vignette, concerning what the normal procedure prescribed or suggested. To this end, the answers to the second dependent variable were recoded from a five-point scale to a three-point scale, whereby the two extremes of the scale were recoded into their two corresponding extremes, and the three middle categories were all recoded into the same single middle category. After that, these recoded scores were subtracted from the corresponding vignette values concerning what the procedures prescribed regarding who should be involved in further recovery steps, which provided the value for the deviation from those procedures. Positive values corresponded with taking on a larger than prescribed share, whereas negative values corresponded with involving others to a larger extent than prescribed.

Appendix 3 contains some examples of the vignettes that were used in studies 4 and 5, translated from Dutch into English.

4.4.4 Questionnaire for person-related factors

In addition to the vignettes, covering the situation-related factors, I also gave each participant a small questionnaire to fill out, to establish what their scores were on the person-related factors that were proposed in section 4.2, occupational self-efficacy and number of years of relevant work experience. I used the short version of Schyns and Von Collani's (2002) occupational self-efficacy scale, consisting of eight items. I translated the items into Dutch to match the participants' language. The items were all positively formulated statements, for which the participants had to indicate the extent to which they agreed or disagreed with the statement on a five point scale, whereby 1 corresponded with 'I totally disagree', and 5 with 'I totally agree'. Below this list of items I asked them to fill in their number of years of experience in process operations. An English version of the person-related factors questionnaire used in the follow-up studies is included in appendix 4.

The scale reliability as expressed in terms of Cronbach's α (standardized) for the occupational self-efficacy scale used in studies 4 and 5 was 0.80 in study 4, 0.65 in study 5, and 0.75 when combining both datasets. The cut-off value of 0.70 generally used for scale reliability (e.g. Nunnally & Bernstein, 1994) was not reached in study 5, involving a much smaller number of participants than study 4 (N=70 versus N=152, respectively, see next subsection). But as the intention was to combine the two datasets for further analyses, an issue that will be discussed in the beginning of section 4.5, I decided to keep this measure. After all, for the combined dataset the reliability of the occupational self-efficacy scale was acceptable.

4.4.5 Distributing the vignettes and questionnaires

As was pointed out by, for example, Rossi & Anderson (1982), when distributing vignettes, i.e. when creating subsets of all the possible vignettes for each of the participants to respond to, it is important to make sure that the resulting study design is as orthogonal as possible. This means that for every factor included in the vignettes, every level of that factor should be roughly equally represented in the entire set of distributed vignettes. Given the impossibility of certain combinations of factor levels, an entirely orthogonal design was not possible. A Visual Basic program[4] was specifically written to at least strive for the best possible approximation of orthogonality in distributing the vignettes over the participants, to ensure that none of the levels or values of the situation-related factors would be severely overrepresented in the distributed vignettes. Four different vignettes were given to each participant. As a result of the program's selection methods, these subsets of four vignettes could be considered as a random selection from all possible vignettes, presented in a random order. Since for each vignette, as was explained before, the corresponding

[4] I would like to thank Dr. Harry Garst from the Department of Technology Management of Eindhoven University of Technology for writing this program. A copy of the macro that was used can be sent to those readers who are interested in the exact distribution procedure that was followed.

questions had to be answered for four different circumstances under which the described deviation could occur, this resulted in an actual total of 16 different scenarios to which each participant responded. Next to the vignettes, the participants were asked to fill out the person-related factors questionnaire, which was introduced as containing some general questions with regard to their work. Based on the program's calculations, for all of the participants envelopes were made, including the vignettes that were selected for them, and a copy of the questionnaire.

In the plant involved in study 4, the participating operators were given an introduction to the study, in a classroom setting, during a training session that formed part of their normal shift schedule. During such sessions they were free from their normal process operations and control duties. To encourage the participants to mark the answers corresponding with what they would do in practice as opposed to what they think they should do (as these two might be different) when confronted with the situation described in a vignette, they were informed that there were no right and wrong answers, that I was looking for answers corresponding with what they actually would do, that participation would be anonymous and on a voluntary basis, and that no individual answers would be communicated to their supervisors. I also explained to the participants that based on the results both they, their organization, and I could learn more about how recovery processes work and thus learn how to better support them. During the introduction the envelopes were handed out and after the introduction the operators were given the time to complete the vignette questions and the questionnaire. Completed vignettes and questionnaires were put back in the envelopes and handed in. In total, five of such sessions were attended, one for each team of operators of the plant. All of the five teams of operators participated, resulting in a total of 152 participants, who completed a total of 2432 scenarios, of which 2189 (90.0%) had no missing answers and were used for further analysis.

In the plant involved in study 5, the participating operators were given an introduction to the study in their own control room, during their daily production meeting, while they were on their normal shift duties. They were given the same information as the participants in study 4 about the purpose of the study and the conditions of participating. During the introduction the envelopes were handed out and the operators were asked to complete the vignette questions and questionnaire as soon as they would have some spare time later during their shift (which they would normally have in an afternoon shift). Afterwards they could put everything back in the envelope, seal it and return it via a mailbox that was specifically placed in the control room for this purpose. In total, five daily production meetings were attended, one for each team of operators of the plant. Here, too, all of the five teams of operators participated, resulting in a total of 70 participants, who completed a total of 1108 scenarios, of which 1048 (94.6%) had no missing answers and were used for further analysis.

In both plants, it took the participants anywhere between 10 and 30 minutes to answer all the questions. Operators were encouraged to report (verbally or in writing) any difficulties they had in relating to the described deviations, so that the appropriate changes could have been made before the next session, but luckily no such difficulties were experienced.

4.4.6 Data analysis

The choice for a vignette approach resulted in a dataset with multiple levels, that is, a hierarchical nature: the vignettes were nested within the participants in the studies. The design of the studies was also incomplete, as each participant only responded to a subset of all the vignettes, which is a common characteristic of vignette studies. To a certain extent, even though in the distribution of the vignettes a reasonably successful attempt was made to keep this to a minimum, the design was even unbalanced, as some levels of some of the factors were either slightly underrepresented or overrepresented in the dataset. As a result of the nested character of the dataset, some of the variance in responses was likely to be the result of differences in person-related factors, and not merely only of the situation-related factors built into the scenarios (e.g. Snijders & Bosker, 1999; Hox et al., 1991). As Hox et al. (1991) argued, disaggregating the data, treating each scenario as an independent observation, randomly sampled, whereby the corresponding values for the person level variables would be copied across all scenarios to which a participant responded, would ignore this nested structure. Such an approach would therefore violate the assumption that the error terms in the regression equations for the dependent variables are uncorrelated. This approach could lead to finding significant effects that are in fact not significant. Aggregating the data, on the other hand, that is collapsing it over respondents, would make analysis of the effects of the situation-related variables impossible. Separation of the respondent- and scenario levels would be another option (Hox et al., 1991), whereby separate regression equations would be computed for each respondent, but the small number of observations within one respondent would probably lead to unstable estimates. Given the nature of the dataset, the best option was to use an approach that has gained an increasing amount of attention over the last years in the social sciences domain: that of multilevel analysis, using hierarchical linear models (e.g. Snijders & Bosker, 1999).

4.4.7 Multilevel analysis and model building

In multilevel analysis, all aggregation levels present in the dataset are taken into account simultaneously (e.g. Snijders & Bosker, 1999). The lowest aggregation level is referred to as level one, and one aggregation level higher as level two, and so on. In the dataset that resulted from the vignette studies, the scenarios, that is, the situation-related variables, formed level one, and the participants, that is, the person-related variables, formed level two. Multilevel models, in contrast with regular, OLS regression equations, include error terms at each aggregation level.

Multilevel model building involves a number of steps, whereby the model that is estimated is gradually made more complicated by including more independent variables in the model. The modeling process is guided by the hypotheses regarding the dataset, and starts with an intercept-only or empty model, based on which the necessity to take the nested or multilevel character of the data into account can be established. To this end, both the intraclass correlation coefficient (ICC) and the design effect can be calculated (Snijders & Bosker, 1999, pp. 16-26). The ICC is defined as the proportion of variance accounted for by the variables at the highest aggregation level. The design effect is calculated by 1+ (average cluster size-1)*ICC, whereby the cluster size is the average number of observations per highest level unit or cluster, in this case the number of scenarios per respondent. If its value is greater than 2, the nested character of the data should definitely be taken into account during model estimation, i.e. a multilevel approach should be used (see Muthen, http://www.statmodel.com/discussion/messages/12/18.html?SaturdayApril820000848am.

In the following steps of the model building process, first the independent variables at the lowest aggregation level are added to the model, to test if their effects on the dependent variable are significant (i.e. the corresponding regression coefficients are significant). Then these variables are made random at the highest level, introducing a random part of the model, with variances for each of the variables. Their regression coefficients may vary significantly between participants, and this step is a test to establish if this is the case, by looking if any of variances of the level one variables are significant. After that, the independent variables at the highest aggregation level are added, to test if these have significant main effects. Then, if there are hypotheses or expectations to warrant this, cross-level interaction terms can be added. As was mentioned in section 4.1, in the studies described here, such expectations existed. The purpose of this last step is to test whether any of the level two variables significantly influences any of the relationships between a level one variable and the dependent variable.

To test if a step taken in the model building process has led to a significantly improved model, i.e. a better fit of the model, after each consecutive model building step, deviance statistics can be computed. The difference between the new model's deviance and the deviance of the model that resulted from the previous step, Δ deviance, has a χ^2 distribution under the null hypothesis that the more detailed model does not predict significantly better than the smaller model, whereby the associated degrees of freedom, Δ df, correspond with the number of added parameters (Snijders & Bosker, 1999).

For conducting the multilevel analyses, I used the software package MLwiN 1.2[5] (Rasbash, Browne, Healy, Cameron & Charlton, 2002).

4.5 Findings, discussion and conclusions

4.5.1 Some statistics

The means and standard deviations for the situation-related factors, the person-related factors and the three dependent variables are presented in table 27.

The correlation between occupational self-efficacy and the number of years of relevant work experience was 0.184 (p=0.030) in study 4, 0.227(p=0.065, i.e. not significant at $p<0.05$) in study 5, and 0.190 ($p<0.001$) when combining both datasets.

[5] I owe many thanks to Dr. Ad de Jong from the Department of Technology Management of Eindhoven University of Technology for showing me, with great patience, how to use this package and make the most of the data I had collected.

Table 27 Means and standard deviations for all variables, for each of the studies separately and both studies combined

	study 4		study 5		combined	
variable	*mean*	*SD*	*mean*	*SD*	*mean*	*SD*
normal procedure[a]	1.085	0.896	0.973	0.818	1.049	0.873
knowledge & expertise[a]	0.897	0.304	0.891	0.312	0.895	0.307
tools & equipment[a]	0.651	0.477	0.699	0.459	0.666	0.472
urgency[a]	0.474	0.499	0.535	0.499	0.494	0.500
importance[a]	0.542	0.498	0.518	0.500	0.534	0.499
time required[a]	0.652	0.476	0.621	0.485	0.642	0.479
other tasks[a]	0.501	0.500	0.505	0.500	0.502	0.500
availability of others[a]	0.500	0.500	0.501	0.500	0.500	0.500
occupational self-efficacy[b]	3.930	0.411	3.887	0.359	3.918	0.395
number of years of relevant work experience[b]	17.811	9.749	19.227	7.184	18.270	9.021
moment when first recovery step is started after detection of a deviation[a]	1.440	0.635	1.419	0.614	1.433	0.628
extent to which people who detect a deviation would involve others in further recovery[a]	2.533	1.431	2.650	1.452	2.571	1.439
deviation from what normal procedures or routines prescribe or suggest with regard to who should be involved in recovery steps after detection (positive = more self, negative = more others)[a]	0.275	0.936	0.120	0.878	0.225	0.920

[a] *N*=2189, 1048, and 3237 respectively for study 4, 5, and both studies combined, based on only those scenarios where no answers were missing

[b] *N*=137, 66, and 203 respectively for study 4, 5, and both studies combined, based on those participants with no missing answers

4.5.2 Differences between the data from studies 4 and 5

For each of the three dependent variables, as part of the model building process, I tested if there were any significant differences between the datasets from the two follow-up studies that were serious enough to rule out the possibility that the same model for that dependent variable would apply for both organizations. This was done by adding an organizational context dummy variable revealing the origin of the data: this dummy was assigned the value 0 for data from the plant which participated in study 4, and 1 for data from the plant which participated study 5. The existence of significant interactions of this context dummy with the situation-related factors, i.e. scenario level variables (8 possible interactions), or with the person-related factors, i.e. person level variables (another 2 possible interactions), would indicate a difference between the two plants in the relationship of the independent variables involved with the dependent variable.

For the first dependent variable, the priority given to recovery steps, operationalized as the moment when the first recovery step after detection would start, an interaction of the context dummy was found with the availability of the necessary knowledge and expertise ($p<0.05$). The unstandardized regression coefficient for the relationship between availability of the necessary knowledge and expertise and the moment when recovery steps would start was 0.162 for the plant involved in study 4, and 0.162-0.292=-0.130 for the plant involved in study 5. Furthermore, the intercept for the plant involved in study 5 was 0.345 higher than for the plant involved in study 4 (this was the regression coefficient of the context dummy). This indicated that the effect of having the necessary knowledge and expertise available differed slightly between the two organizations. In the plant involved in study 4 the recovery steps tended to start sooner (0.345 scale units) after detection than in the plant involved in study 5, when those detecting the deviation did not have the required knowledge and expertise. However, in cases when those involved in the detection had the knowledge and expertise required for the recovery steps, there was hardly any difference between the two plants, recovery steps would start only slightly (0.053 scale units) sooner in the plant involved in study 4 than in the plant involved in study 5. Since for only one out of 10 factors included in the studies there was a difference between the plants, and on a three point scale, which was used for the dependent variable, this difference in approach between the two organizations was only minor, I decided to ignore this specific difference in further model building.

Regarding the second dependent variable, the distribution of the recovery tasks, operationalized as the extent to which others would be involved in recovery steps, an interaction of the context dummy was found with the amount of time required to perform the necessary recovery steps ($p<0.05$). The unstandardized regression coefficient for the relationship between the time required for recovery and the extent to which others would be involved in the recovery steps was 0.005 and not significant for the plant involved in study 4, and 2.94 for the plant involved in study 5. As the coefficient for the context dummy itself was not significant, there were no differences with regard to the intercept between the plants. This meant that only in the plant involved in study 5 there was an effect of the time required on the extent to which others would be involved in recovery, which was, in line with hypothesis *2f*, a positive effect. Since the results for the plant involved in study 5 at least were not opposite to what was hypothesized, and for only one out of 10 factors included in the studies there was a difference between the plants, I decided to ignore this specific difference between the two organizations in further model building in favor of a simpler, generally applicable model.

Finally, with regard to the third dependent variable, the deviation from what the normal procedures or routines prescribed or suggested with regard to who should be involved in the recovery steps that follow detection, an interaction of the context dummy was found with occupational self-efficacy ($p<0.05$). The unstandardized regression coefficient for the relationship between occupational self-efficacy of a person and the deviation from what the normal procedures or routines prescribed or suggested regarding who should be involved in the recovery steps that follow detection was 0.023 and not significant for the plant involved in study 4, and 0.182 for the plant involved in study 5, with the intercept for the plant involved in study 5 being 0.112 lower than for the plant involved in study 4. This indicated that in the plant involved in study 5 there was a larger tendency towards involving more others than suggested than in the plant involved in study 4, which decreased for people with higher occupational self-efficacy scores than average, and which increased for people with lower than average occupational self-efficacy scores. So again, only for the plant involved in study 5 there was an effect of occupational self-efficacy on the dependent variable. Also in this case, I decided to ignore this specific difference between the two organizations in further model building in favor of a simpler, generally applicable model. After all, here, too, for only one out of 10 factors included in the studies there was a difference between the two plants. Furthermore, recall from subsection 4.4.4 that the scale reliability for occupational self-efficacy in study 5 was just below the commonly accepted cut-off value of 0.70, so anyway, care should be taken not to draw any major conclusions from the difference between the two plants that was identified here.

As a result of the considerations discussed above, for further model building, none of the interactions with the organizational context dummy were kept in the models. To control for its potential main effects, however, the context dummy itself was left in the models, at the highest of the two aggregation levels that were distinguished, based on its nature. As it turned out, in none of the models its effect was significant.

4.5.3 Findings and discussion regarding priority given to recovery

Table 28 shows the results of the multilevel analysis and model building process for the priority given to recovery, operationalized in the follow-up studies as the moment when the first recovery step starts after a deviation is detected. I followed the steps described in subsection 4.4.7:

- to test hypothesis 1, regarding the significance and direction of the effects of each of the situation-related factors, i.e. scenario level variables, on the priority given to recovery;
- to answer the research question A1, dealing with the main effect of the person-related factor occupational self-efficacy on the priority given to recovery;
- to answer the research question B1, dealing with the main effect of the person-related factor experience on the priority given to recovery;
- to answer the first part of research question A4, dealing with the possibility of significant interactive effects of the situation-related factors and the person-related factor occupational self-efficacy on the priority given to recovery; and
- to answer the first part of research question B4, dealing with the possibility of significant interactive effects of the situation-related factors and the person-related factor experience on the priority given to recovery.

The corresponding findings are discussed in the remainder of this subsection.

Table 28 Multilevel analysis results for moment when first recovery step after detection starts, with unstandardized regression coefficients for all independent variables and interactions in fixed part, variances for scenario level variables in random part
(standard errors between parentheses)

recovery starts when	empty model	model 1	model 2	model 3	model 4
fixed part					
intercept	1.433(0.022)**	1.632(0.049)**	1.700(0.071)**	1.723(0.071)**	1.713(0.072)**
scenario level coefficients					
normal procedure		0.113(0.013)**	0.070(0.019)**	0.070(0.019)**	0.069(0.019)**
knowledge & expertise		0.069(0.035)**	0.042(0.054)	0.036(0.053)	0.054(0.053)
tools & equipment		-0.129(0.022)**	-0.152(0.036)**	-0.151(0.036)**	-0.156(0.036)**
urgency		-0.341(0.020)**	-0.358(0.034)*	-0.359(0.034)**	-0.360(0.034)**
importance		-0.334(0.020)**	-0.304(0.039)**	-0.304(0.039)**	-0.300(0.039)**
time required		0.014(0.020)	0.006(0.031)	0.005(0.031)	0.002(0.031)
other tasks		0.184(0.018)**	0.186(0.021)**	0.187(0.021)**	0.186(0.020)**
availability of others		-0.098(0.018)**	-0.097(0.013)**	-0.097(0.013)**	-0.097(0.013)**
person level coefficients					
occupational self-efficacy (ose)[a]				-0.041(0.041)	0.231(0.187)
years relevant experience (exp)[a]				0.003(0.002)	0.003(0.008)
organizational context dummy				-0.045(0.035)	-0.044(0.035)
cross-level interactions					
ose*normal procedure					-0.055(0.051)
ose*knowledge & expertise					-0.185(0.132)
ose*tools & equipment					0.097(0.104)
ose*urgency					-0.102(0.085)
ose*importance					-0.136(0.101)
ose*time required					-0.028(0.078)
ose*other tasks					0.105(0.052)**
ose*availability of others					-0.010(0.033)
exp*normal procedure					0.002(0.002)
exp*knowledge & expertise					0.004(0.006)
exp*tools & equipment					-0.005(0.004)
exp*urgency					0.002(0.004)
exp*importance					-0.004(0.004)
exp*time required					0.001(0.004)
exp*other tasks					-0.005(0.002)**
exp*availability of others					0.004(0.001)**

recovery starts when	empty model	model 1	model 2	model 3	model 4
random part					
variance at participants level	0.082(0.010)**	0.056(0.007)**	0.283(0.093)**	0.272(0.086)**	0.278(0.092)**
variance normal procedure			0.025(0.007)**	0.024(0.006)**	0.024(0.006)**
variance knowledge & expertise			0.117(0.045)**	0.104(0.040)**	0.110(0.043)**
variance tools & equipment			0.107(0.024)**	0.106(0.024)**	0.103(0.023)**
variance urgency			0.133(0.023)**	0.136(0.023)**	0.127(0.022)**
variance importance			0.214(0.031)**	0.215(0.031)**	0.215(0.031)**
variance time required			0.089(0.019)**	0.090(0.019)**	0.091(0.019)**
variance other tasks			0.052(0.009)**	0.052(0.009)**	0.049(0.008)**
variance availability of others			0.001(0.004)	0.001(0.004)	0.000(0.004)
variance at scenario level	0.312(0.008)**	0.248(0.006)**	0.135(0.004)**	0.135(0.004)**	0.135(0.004)**
deviance	5752.210	4986.468	4156.506	4153.125	4125.931
Δ deviance		765.742***	829.962***	3.381	27.194**
Δ df		8	44	3	16
$R_1^{2\ b}$		0.228	0.228	0.234	0.292
$R_2^{2\ c}$		0.296	0.296	0.315	0.320
ICC	0.208				
design effect	4.122				

* $p<0.1$
** $p<0.05$
*** $p<0.01$
[a] centered around grand mean (i.e. mean for entire dataset)
[b] explained variance at level one, the scenario level
[c] explained variance at level two, the person level

The design effect, which was far greater than 2 (cf. the description of multilevel analysis in subsection 4.4.7), clearly demonstrated the nested character of the data and the need for a multilevel approach in data analysis. As can be seen from the significant Δ deviance, step 1, which produced model 1, resulted in a significantly better model fit than the empty model, indicating that adding the scenario level variables led to an improved model. Step 2 where the scenario level effects were made random at person level, producing model 2, again resulted in a significantly better model. In step 3 where the person level variables main effects were added, producing model 3, the model fit did not significantly improve, Δ deviance was not significant. In step 4, however, where the cross-level interactions of occupational self-efficacy and experience with the scenario level variables were included, producing model 4, the model fit improved significantly again. And in order to include the cross-level interactions, the person level variables have to be included as main effects first. Therefore, to get to the best-fitting, most detailed model, the third model building step where the model fit did not improve significantly had to be taken for granted. The most detailed model, model 4, was used to test the hypotheses and answer the research questions regarding the priority given to recovery.

In multilevel analysis, the proportion of explained variance is calculated somewhat differently than in traditional multiple linear regression (Snijders & Bosker, 1999, pp. 99-109). Two levels can be distinguished for the explained variance, not only the lowest aggregation level the data, in this case corresponding with the scenario level, but also the highest level, corresponding with the person level. The explained variance at level one, the scenario level, is defined as the proportional reduction of error (compared to the empty model) for predicting and individual outcome. The explained variance at level two, the person level, is defined as the proportional reduction of error for predicting a group mean (here a mean for a participant). Explained variance of models with random regression coefficients (from model 2 onwards) can be approximated by the explained variance of the same models with all regression coefficients fixed (Snijders & Bosker, 1999, pp. 104-105). The explained variance at the scenario level of the final model, model 4, was 29.2%. The explained variance at the person level of this model was 32.0%. These percentages of explained variance are certainly reasonable, compared to findings from other researchers who used a multilevel approach for analyzing their data. For example, for one of the multilevel model examples they used in their book, based on a dataset from a study in the domain of social sciences, Snijders & Bosker (1999) found percentages of 44% and 58% respectively for explained variance at level one and level two of and considered these percentages as high. Furthermore, Griffin, Mathieu, and Jacobs (2001) reported percentages of 14.8% and 32.7% respectively for explained variance at level one and level two of their best fitting model for a study in the domain of work and organizational psychology.

As can be seen from the regression coefficients for model 4, for the scenario level variables in the ‘fixed part’ section of table 28, most parts of hypothesis 1 about the factors influencing priority given to recovery were confirmed. Only part *f)*, concerning the effect of time required for recovery tasks, and part *b)*, concerning the effect of availability of the required knowledge and expertise, were not confirmed, since the corresponding the regression coefficients were not significant. The regression coefficients of the other six scenario level variables were all significant into the hypothesized directions. The findings indicated that highest priority was given to urgent and important tasks, where the extent to which others should be involved according to the normal procedure was low, where tools and equipment were available, where the people detecting the deviation were not busy with other urgent and important tasks, and where others who could assist were available. To enable a comparison of the strength of their effects, the regression coefficients for these six scenario level variables were standardized, by multiplying the unstandardized coefficients as given in the table above with quotient of the standard deviation of the corresponding variable and the standard deviation of the dependent variable in the model. As expected, the effects of urgency (standardized regression coefficient –0.287) and importance (standardized regression coefficient –0.238) turned out to be the strongest, with urgency indeed outweighing importance, which was in line with other researchers' predictions (e.g. Koch & Kleinmann, 2002) and findings (Claessens et al., 2003). The effect of the availability of others on prioritization was rather small (standardized regression coefficient –0.077), apparently this was not often reason enough to delay the start of a recovery process. The effect of the normal procedures regarding who should be involved in the recovery steps after detection was also quite small (standardized regression coefficient 0.096). Having to involve others did not delay the rest of the recovery process very much compared to situations where people are allowed to perform all steps alone, especially as long as they were allowed to stay involved to some extent. Participants in the studies commented that if the recovery steps are really urgent and important, and additional people are needed who are not currently present, these can usually be on the spot within very little time, so that delays are kept as small as possible. The effect of not having the necessary tools and equipment was a little stronger (standardized regression coefficient –0.117), which could indicate that in some cases it may not be so easy to get these on the spot. The effect of having other urgent and important tasks was again a little stronger (standardized regression coefficient 0.148), obviously these tasks could compete for people's attention. In line with the findings from the initial studies regarding task prioritization, the findings described here again underlined the importance of ensuring that the people within the organization have a complete and correct understanding of the potential consequences of possible deviations, both with regard to type and severity, and with regard to how soon they can occur, so that prioritization will be based on the correct information. Furthermore, it was reassuring to see that other factors that could delay the start of further recovery steps in fact had a much smaller effect on establishing priorities than the urgency and importance of recovery.

As for the factors with non-significant regression coefficients, the time required for recovery steps, corresponding with part *f)* of hypothesis 1, apparently did not influence the moment when the first recovery step was started. This indicated that where recovery tasks are concerned, people did not let the thought that these tasks would take a long time to complete deter them from getting started. This was a very positive finding indeed, as especially for recovery tasks the consequences of not performing those in time can be disastrous. As for the availability of the required knowledge and expertise, corresponding with part *b)* of hypothesis 1, in the first model this effect had the opposite direction than what was expected. When those involved in detection had the knowledge and expertise required for the recovery steps, this led to a slightly later start of the recovery steps. An explanation for this finding could be that if people did not know how to perform the required recovery step, they tended to build in some extra time for these steps by getting them started early. This effect, however, was no longer significant as soon as the scenario level effects were made random at person level, so from model 2 onwards, and remained not significant after the inclusion of the cross-level interactions. This finding could indicate that people in fact knew very well exactly which priority the recovery tasks deserved, that is, when they should be started, without necessarily having the knowledge and expertise needed to be able to perform these tasks. So, there were no delays as a result of having to find others, and neither was extra time built in, by getting them started sooner, for tasks that those involved in detection could not perform personally. This served as an indication of how the operators' general plant and process knowledge, i.e. regarding what can lead where, can be beneficial for recovery.

As for research questions A1 and B1, focusing on the main effects of occupational self-efficacy and experience, respectively, the non-significant regression coefficients for these variables in models 3 and 4 indicated that there were no main effects of occupational self-efficacy and experience on the moment when the first recovery steps after detection started. While there obviously was some variance in the intercept at person level, this could not be explained via the main effects of the variables that were included at this level in the studies. An explanation for this finding could be the small variance between the participants with regard to their scores on the person level variables. If the distribution across the answer categories for the occupational self efficacy scale would have been uniform, the standard deviation would have been 2, whereas it was only 0.396. Operators resembled each other to a great extent with regard to occupational self-efficacy and all had relatively high scores on this scale. In fact, their self-efficacy may have been one of the reasons why they chose their job, or were chosen for it, and stuck with that choice in the first place. Their self-efficacy may also have been the result of good training in this line of work. To handle the responsibilities and demands attached to controlling high risk processes, operators have to believe in their own abilities. The variance between the participants with regard to the number of years of relevant work experience was relatively small too, and on top of that, especially the group of operators with little experience was underrepresented in the dataset. This lack of variance in experience and the high mean for number of years of experience are quite common for the industry, and is basically the result of the organizational restructuring processes that have been going on in the industry during the last decades. Furthermore, possibly the occupational self-efficacy scale was not task-specific enough to identify differences between the participants that were relevant specifically for situations where recovery is needed. In addition, possibly other person level factors that were not measured in the follow-up studies, such as personal preferences with regard to which tasks one finds most appealing and with regard to task planning, could help to explain person level variance, too.

As can be seen from the ‘random part’ section of table 28, all of the scenario level variables, except the availability of others who could assist, had significant variances when made random at person level (from model 2 onwards), which meant that there were differences between the participants with regard to the regression coefficients for these variables. A common method to explain these variances is by looking at cross-level interactions between the lowest, i.e. scenario level variables, and the variables at a higher aggregation level, in this case the person or participant level. Cross-level interactions with the availability of others who could assist were also explored, since even if a lower level variable is not random at a higher aggregation level, such interactions can still exist (Snijders & Bosker, 1999). As can be seen from the regression coefficients for the cross-level interactions in model 4, the only significant cross-level interaction between occupational self-efficacy and the scenario level variables was the interaction with the presence of other urgent and important tasks. This answered the first part of research question A4, which aimed at identifying interactive effects of scenario level variables with occupational self-efficacy. The regression coefficients for the cross-level interactions in model 4 also show that the only significant cross-level interactions between experience and the scenario level variables were the interactions with the presence of other urgent and important tasks, and with the availability of others who can assist. This answered the first part of research question B4, which aimed at identifying interactive effects of scenario level variables with experience.

To interpret interactions whereby one of the two variables involved (here the situation-related factor) can only take two values, two regression equations have to be compared: one equation in which one of the two possible values is substituted in the place of the variable concerned, with another equation in which the other possible value is substituted in the same place. The difference between the regression equations for the moment when the first recovery steps after detection started, for a situation with, and a situation without other already ongoing urgent and important tasks, was 0.186+0.105*ose (based on unstandardized regression coefficients, whereby ose is the score on occupational self-efficacy, centered around the grand mean for the entire dataset). So, this positive difference was larger for people with a higher than average score on occupational self-efficacy, and smaller for people with a lower than average score on occupational self-efficacy. For the difference to become negative, people would need to have a score of 1.8 lower than the average score on occupational self-efficacy. The dataset did not contain any such cases. These findings could be an indication that while more efficacious people would not hesitate to delay recovery steps in favor of other urgent and important tasks, less efficacious people do not have the confidence to do this to the same extent, they may prefer other ways of dealing with the additional workload.

The difference between the regression equations for the moment when the first recovery steps after detection started, for a situation with, and a situation without other urgent and important tasks, was 0.186-0.005*exp (based on unstandardized regression coefficients, whereby exp is the number of years of experience, centered around the grand mean for the entire dataset). So, this positive difference was slightly larger for people with less than the average number of years of experience and slightly smaller for people with more than the average number of years of experience. For the difference to become negative, people would need to have roughly 37.3 years more experience than average. The dataset did not contain any such cases. These findings could indicate that more experienced people do not allow other tasks to delay them as long as less experienced people do, or they may actually be able to deal with those tasks more quickly.

The difference between the regression equations for the moment when the first recovery steps after detection started, for a situation where others who could assist were available, and a situation where no such others were available, was –0.097+0.004*exp (based on unstandardized regression coefficients, whereby exp is the number of years of experience, centered around the grand mean for the entire dataset). So, this negative difference was slightly larger for people with less than the average number of years of experience, and slightly smaller for people with more than average experience. For the difference to become positive, people would need to have roughly 24.3 years more experience than average. The dataset did not contain any such cases. These findings served as an indication that less experienced people depended more on the assistance of others than more experienced people to get the recovery steps started.

4.5.4 Findings and discussion regarding distribution of recovery tasks

Table 29 shows the results of the multilevel analysis and model building process for the distribution of recovery tasks, operationalized in the follow-up studies as the extent to which people who detect a deviation decide to involve others in the recovery steps that follow. I followed the steps described in subsection 4.4.7:

- to test hypothesis 2, regarding the significance and direction of the effects of each of the situation-related factors, i.e. scenario level variables, on the distribution of recovery tasks;
- to answer the research question A2, dealing with the main effect of the person-related factor occupational self-efficacy on the distribution of recovery tasks;

- to answer the research question B2, dealing with the main effect of the person-related factor experience on the distribution of recovery tasks;
- to answer the second part of research question A4, dealing with the possibility of significant interactive effects of the situation-related factors and the person-related factor occupational self-efficacy on the distribution of recovery tasks; and
- to answer the second part of research question B4, dealing with the possibility of significant interactive effects of the situation-related factors and the person-related factor experience on the distribution of recovery tasks.

The corresponding findings are discussed in the remainder of this subsection.

The design effect was again far greater than 2, demonstrating the nested character of the data and the need for a multilevel approach in data analysis. As can be seen from the significant Δ deviance, step 1, which produced model 1, resulted in a significantly better model fit than the empty model, indicating that adding the scenario level variables led to an improved model. Step 2 where the scenario level effects were made random at person level, producing model 2, again resulted in a significantly better model. In step 3 where the person level variables main effects were added, producing model 3, Δ deviance was not significant, so the model fit did not significantly improve. However, including the cross-level interactions of occupational self-efficacy and experience with the scenario level variables in step 4, producing model 4, resulted again in a significantly better model than the previous model. In order to include these cross-level interactions, the person level variables have to be included as main effects first, so to get to the best-fitting, most detailed model, the third model building step where the model fit did not improve significantly had to be accepted. The most detailed model, model 4, was used to test the hypotheses and answer the research questions regarding the distribution of recovery tasks.

Similar as in the previous subsection, for the models presented in this subsection the proportions of explained variance, both at the scenario level and at the person level, were calculated based on formulae provided by Snijders and Bosker (1999, pp. 104-105). The explained variance at the scenario level of the final model, model 4, was 31.1%. The explained variance at the person level of this model was 30.0%.

Table 29 Multilevel analysis results for extent to which those detecting a deviation involve others in further recovery, with unstandardized regression coefficients for all independent variables and interactions in fixed part, variances for scenario level variables in random part (standard errors between parentheses)

recovery by whom	empty model	model 1	model 2	model 3	model 4
fixed part					
intercept	2.573(0.046)**	2.466(0.108)**	2.313(0.0147)**	2.262(0.149)**	2.309(0.148)**
scenario level coefficients					
normal procedure		0.389(0.029)**	0.427(0.041)**	0.426(0.041)**	0.422(0.041)**
knowledge & expertise		-0.817(0.078)**	-0.688(0.125)**	-0.672(0.125)**	-0.686(0.123)**
tools & equipment		-0.187(0.049)**	-0.185(0.072)**	-0.187(0.072)**	-0.216(0.072)**
urgency		-0.274(0.044)**	-0.294(0.060)**	-0.297(0.060)**	-0.295(0.059)**
importance		-0.252(0.044)**	-0.273(0.069)**	-0.268(0.069)**	-0.271(0.068)**
time required		0.111(0.045)**	0.109(0.064)*	0.112(0.065)**	0.107(0.064)*
other tasks		0.647(0.039)**	0.647(0.040)**	0.647(0.040)**	0.646(0.040)**
availability of others		0.853(0.039)**	0.854(0.051)**	0.854(0.051)**	0.858(0.050)**
person level coefficients					
occupational self-efficacy (ose)[a]				-0.045(0.084)	0.800(0.378)**
years relevant experience (exp)[a]				0.005(0.004)	-0.014(0.017)
organizational context dummy				0.106(0.071)	0.108(0.071)
cross-level interactions					
ose*normal procedure					-0.047(0.108)
ose*knowledge & expertise					-0.772(0.304)**
ose*tools & equipment					0.195(0.205)
ose*urgency					-0.228(0.150)
ose*importance					-0.018(0.174)
ose*time required					-0.128(0.165)
ose*other tasks					0.162(0.103)
ose*availability of others					-0.350(0.129)**
exp*normal procedure					0.001(0.005)
exp*knowledge & expertise					0.036(0.014)**
exp*tools & equipment					-0.014(0.008)*
exp*urgency					0.003(0.007)
exp*importance					-0.010(0.008)
exp*time required					0.006(0.007)
exp*other tasks					-0.005(0.005)
exp*availability of others					-0.002(0.006)

recovery by whom	empty model	model 1	model 2	model 3	model 4
random part					
variance at participants level	0.322(0.043)**	0.244(0.032)**	1.136(0.419)**	1.142(0.421)**	1.198(0.381)**
variance normal procedure			0.141(0.032)**	0.141(0.032)**	0.137(0.031)**
variance knowledge & expertise			0.862(0.269)**	0.847(0.286)**	0.877(0.264)**
variance tools & equipment			0.371(0.094)**	0.388(0.096)**	0.380(0.094)**
variance urgency			0.282(0.074)**	0.285(0.074)**	0.263(0.071)**
variance importance			0.523(0.095)**	0.526(0.096)**	0.514(0.094)**
variance time required			0.370(0.083)**	0.371(0.083)**	0.376(0.083)**
variance other tasks			0.148(0.033)**	0.148(0.033)**	0.144(0.033)**
variance availability of others			0.352(0.053)**	0.352(0.053)**	0.332(0.051)**
variance at scenario level	1.746(0.045)**	1.215(0.031)**	0.721(0.022)**	0.721(0.022)**	0.7210.022)**
deviance	11270.550	10109.130	9494.547	9490.438	9457.677
Δ deviance		1161.42***	614.583***	4.109	32.761***
Δ df		8	44	3	16
$R_1^{2\ b}$		0.294	0.294	0.299	0.311
$R_2^{2\ c}$		0.258	0.258	0.281	0.300
ICC	0.156				
design effect	3.336				

* *p*<0.1
** *p*<0.05
*** *p*<0.01
[a] centered around grand mean (i.e. mean for entire dataset)
[b] explained variance at level one, the scenario level
[c] explained variance at level two, the person level

As can be seen from the regression coefficients for model 4, for the scenario level variables in the ‘fixed part’ section of table 29, all parts of hypothesis 2 about the factors influencing the distribution of recovery tasks were confirmed. The regression coefficients of the all of the eight scenario level variables were all significant into the hypothesized directions. The findings indicated that people tended to involve more others in cases where procedures suggested to do so, where they did not have the required knowledge and expertise, nor the necessary tools and equipment, and for less urgent and important recovery steps, where greater amounts of time were required for the recovery steps, and in cases where the other tasks on which they were already working were more urgent and important, and as long as others who could assist were at least available. To enable a comparison of the strength of their effects, the regression coefficients for these eight scenario level variables were standardized, by multiplying the unstandardized coefficients as given in the table above with quotient of the standard deviation of the corresponding variable and the standard deviation of the dependent variable in the model. The involvement of others was expected to be dictated by what the normal procedures or routines suggested in this regard, which was confirmed by the substantial, positive value of the corresponding standardized regression coefficient (0.256). Some of the other scenario level variables also had quite a large effect on the dependent variable, most notably the availability of others (standardized regression coefficient 0.298). This made sense, as absence of such others made it impossible to get them involved right away. The presence of other urgent and important tasks also had a substantial effect (standardized regression coefficient 0.225), which could indicate that operators were not inclined to abandon such tasks if involving others in recovery was an option. Not having the required knowledge and expertise had quite a substantial effect on involving others, too (standardized regression coefficient –0.146), demonstrating that people did not easily take on tasks they were not sure how to handle. Furthermore, both the urgency and importance of the recovery steps played a role, even though their effect was weaker than the effects mentioned above (standardized regression coefficients respectively –0.103 and –0.094), and also weaker than the effect the same two factors had on prioritization of the recovery steps. The effect of the availability of the necessary tools and equipment was again a little weaker (standardized regression coefficient –0.071), but apparently not having these was still a reason to involve others, either because they were the ones who had these tools and equipment, or they simply could have been the ones on duty as soon as the required tools and equipment arrived. The effect of the time required for recovery on who was involved in the recovery steps was the smallest of all the effects (standardized regression coefficient –0.036). This could be an indication that the operators did not specifically push recovery tasks involving more work to others while not minding to take on short recovery tasks on their own – at least this tendency was not that great. These findings underlined the importance of especially the availability of others on who was involved in the recovery steps following detection. In light thereof, organizations may want to ensure that for recovery tasks for which the involvement of others is critical, contacting these others and getting them to where they are needed is made as easy as possible. This may sound trivial, but as a result of the restructuring processes that have taken place in many organizations in recent years whereby the amount of employees has been reduced to a minimum, in some organizations it is not so self-evident that such others are easily available when they are needed.

Regarding research question A2, focusing on the main effect of occupational self-efficacy on the extent to which others would be involved in the recovery steps following detection, the corresponding regression coefficient became significant in model 4, as soon as the cross-level interactions were also taken into account. The main effect of occupational self-efficacy on the involvement of others in recovery steps that followed detection was positive, indicating that people with higher than average self-efficacy were more inclined to involve others in recovery than people with lower than average self-efficacy. This finding seemed a little surprising, as one would expect that more efficacious people would more readily take on larger shares of the recovery steps. Anyway, given the presence of interactions of this variable with scenario level variables, the main effect should be considered in light of these interactions, addressed in the next paragraph. As for research question B2, focusing on the main effect of experience on the extent to which others would be involved in the recovery steps following detection, the non-significant regression coefficient for number of years of relevant work experience both in models 3 and 4 indicated that there was no such main effect. As was argued already in the discussion of the findings regarding research question B1 in the previous subsection, an explanation for this finding could be the relative homogeneity of the group of operators who participated in the studies with regard to experience.

As can be seen from the 'random part' section of table 29, all of the scenario level variables had significant variances when made random at person level (from model 2 onwards), which meant that there were differences between the participants with regard to the regression coefficients for these variables. As in the previous subsection, here, too, an attempt was made to explain these differences by looking for significant cross-level interactions between the scenario-related variables and the person-related variables. The regression coefficients for the cross-level interactions in model 4 show that the only significant cross-level interactions between occupational self-efficacy and the scenario level variables were the interactions with the availability of the required knowledge and expertise, and with the availability of others who can assist. This answered the second part of research question A4, which aimed at identifying interactive effects of scenario level variables with occupational self-efficacy. The regression coefficients for the cross-level interactions in model 4 also show that the only significant cross-level interactions between experience and the scenario level variables were the interactions with the availability of the required knowledge and expertise, and with the availability of the necessary tools and equipment. This answered the second part of research question B4, which aimed at identifying interactive effects of scenario level variables with experience.

To interpret these interactions, a similar approach was used as in the previous subsection, that is, two regression equations were compared, one equation with one of the two possible values of the situation-related factor involved filled in in the equation, with another equation in which the other possible value was filled in. The difference between the regression equations for the extent to which others were involved in the recovery steps after detection, for a situation where those who detected the deviation had the required knowledge and expertise, and a situation where they did not have the required knowledge and expertise, was –0.686–0.772*ose (based on unstandardized regression coefficients, whereby ose is the score on occupational self-efficacy, centered around the grand mean for the entire dataset). Add to this the main effect of occupational self-efficacy (unstandardized regression coefficient 0.800), which was significant in this model, and which applied for both equations. Keeping all other factors constant, the regression line for the effect of occupational self-efficacy on the extent to which others would be involved in recovery, for the situation where those who detected a deviation had the required knowledge and expertise, lies below the line for the situation where those who detected a deviation did not have the required knowledge and expertise, and hardly shows an upward trend. The upward trend for the other line is much greater. If a person had the required knowledge and expertise to deal with a detected deviation, this greatly reduced the positive effect of occupational self-efficacy on the extent to which this person decided to involve others in the recovery tasks following detection. More efficacious people were more influenced by having the required knowledge and expertise than their less efficacious coworkers, in their decision regarding the extent to which they would involve others.

The difference between the regression equations for the extent to which others were involved in the recovery steps after detection, for a situation where others who could assist were available, and a situation where no such others were available, was 0.858–0.350*ose (based on unstandardized regression coefficients, whereby ose is the score on occupational self-efficacy, centered around the grand mean for the entire dataset). Add to this the main effect of occupational self-efficacy (unstandardized regression coefficient 0.800), which was significant in this model, and which applied for both equations. Keeping all other factors constant, the regression line for the effect of occupational self-efficacy on the extent to which others would be involved in recovery, for the situation where others who could assist were available, starts above the line for the situation where no such others were available, but shows less of an upward trend. If others who could assist were available, this greatly reduced the positive effect of occupational self-efficacy on the extent to which this person decided to involve others in the recovery tasks following detection. More efficacious people were less influenced by the availability of others in their decision to involve others than their less efficacious coworkers.

The difference between the regression equations for the extent to which others were involved in the recovery steps after detection, for a situation where those who detected the deviation had the required knowledge and expertise, and a situation where they did not have the required knowledge and expertise, was –0.686+0.036*exp (based on unstandardized regression coefficients, whereby exp is the number of years of experience, centered around the grand mean for the entire dataset). So, this negative difference was larger for people with a lower than number of years of experience, and smaller for people with a higher than average number of years of experience, or for them the difference could even become positive. This indicated that based on having the required knowledge and expertise, for people with a lower than average number of years of experience, the extent to which they would involve others in further recovery tasks tended to shift to lower levels than for their more experienced coworkers in the same circumstances. For people with roughly 19.1 or more years more than the average number of years of experience, the direction of the effect of having the required knowledge and expertise on the extent to which people who detected a deviation would involve others in further recovery even changed, towards involving more others. A possible reason for very experienced people to involve more others is to share this experience with others, to transfer part of their experience to these others, whereas less experienced people could feel a greater need to use all the practice they can get in applying their knowledge.

The difference between the regression equations for the extent to which others were involved in the recovery steps after detection, for a situation where those who detected the deviation had the required tools and equipment, and a situation where they did not have the required tools and equipment, was –0.216–0.014*exp (based on unstandardized regression coefficients, whereby exp is the number of years of experience, centered around the grand mean for the entire dataset). So, this negative difference was larger for people with a higher than average number of years of experience, and smaller for people with a lower than average number of years of experience, or for them this difference could even become positive. This indicated that based on having the required tools and equipment, for people with a higher than average number of years of experience, the extent to which they would involve others in further recovery tasks tended to shift to lower levels than for their less experienced coworkers in the same circumstances. For people with roughly 15.5 or more years less than the average number of years of experience, the direction of the effect of having the required knowledge and expertise on the extent to which people who detected a deviation would involve others in further recovery even changed, towards involving more others. Apparently, a basic level of experience was needed before people were inclined to allow the availability of the required tools and equipment to shift the extent to which they would involve others in further recovery towards lower levels.

4.5.5 Findings and discussion concerning deviations from what procedures prescribe regarding who should be involved in recovery tasks

Table 30 shows the results of the multilevel analysis and model building process for the deviation from what the normal procedures or routines prescribe or suggest with regard to who should be involved in the recovery steps that follow detection.

Table 30 Multilevel analysis results for deviation from what normal procedures prescribe regarding who to involve in recovery steps after detection, with unstandardized regression coefficients for all independent variables and interactions in fixed part, variances for scenario level variables in random part (standard errors between parentheses)

procedure deviation (negative=more others, positive=more self)	empty model	model 1	model 2	model 3	model 4
fixed part					
intercept	0.225(0.031)**	-0.651(0.052)**	-0.613(0.074)**	-0.587(0.075)**	-0.599(0.075)**
scenario level coefficients					
normal procedure		0.823(0.014)**	0.810(0.019)**	0.810(0.019)**	0.812(0.019)**
knowledge & expertise		0.270(0.038)**	0.241(0.062)**	0.230(0.062)**	0.230(0.062)**
tools & equipment		0.065(0.024)**	0.077(0.035)**	0.080(0.035)**	0.089(0.035)**
urgency		0.108(0.021)**	0.114(0.027)**	0.117(0.027)**	0.118(0.027)**
importance		0.078(0.021)**	0.085(0.033)**	0.084(0.033)**	0.084(0.033)**
time required		-0.071(0.022)**	-0.082(0.029)	-0.084(0.030)**	-0.080(0.029)**
other tasks		-0.260(0.019)**	-0.260(0.019)**	-0.260(0.019)**	-0.260(0.019)**
availability of others		-0.382(0.019)**	-0.383(0.026)**	-0.383(0.026)**	-0.385(0.025)**
person level coefficients					
occupational self-efficacy (ose)[a]				0.073(0.040)*	-0.296(0.189)
years relevant experience (exp)[a]				-0.004(0.002)**	0.001(0.008)
organizational context dummy				-0.052(0.033)	-0.054(0.033)
cross-level interactions					
ose*normal procedure					0.002(0.049)
ose*knowledge & expertise					0.264(0.151)*
ose*tools & equipment					-0.005(0.100)
ose*urgency					0.086(0.068)
ose*importance					0.007(0.084)
ose*time required					0.032(0.075)
ose*other tasks					-0.058(0.048)
ose*availability of others					0.162(0.065)**
exp*normal procedure					0.001(0.002)
exp*knowledge & expertise					-0.010(0.007)
exp*tools & equipment					0.007(0.004)*
exp*urgency					-0.001(0.003)
exp*importance					0.002(0.004)
exp*time required					-0.004(0.003)
exp*other tasks					0.001(0.002)
exp*availability of others					0.001(0.003)

procedure deviation **(negative=more others, positive=more self)**	**empty model**	**model 1**	**model 2**	**model 3**	**model 4**
random part					
variance at participants level	0.153(0.019)**	0.048(0.007)**	0.298(0.104)**	0.304(0.105)**	0.325(0.104)**
variance normal procedure			0.026(0.007)**	0.026(0.007)**	0.025(0.006)**
variance knowledge & expertise			0.192(0.063)**	0.187(0.062)**	0.194(0.062)**
variance tools & equipment			0.087(0.022)**	0.093(0.023)**	0.095(0.023)**
variance urgency			0.045(0.013)**	0.045(0.015)**	0.043(0.015)**
variance importance			0.113(0.021)**	0.114(0.021)**	0.118(0.022)**
variance time required			0.066(0.017)**	0.067(0.018)**	0.066(0.017)**
variance other tasks			0.025(0.007)**	0.025(0.007)**	0.025(0.007)**
variance availability of others			0.088(0.013)**	0.088(0.013)**	0.083(0.013)**
variance at scenario level	0.693(0.018)**	0.287(0.007)**	0.182(0.006)**	0.182(0.006)**	0.182(0.006)**
deviance	8305.895	5414.406	4890.627	4881.401	4859.767
Δ deviance		2891.489***	523.779***	9.226**	21.634 (p<0.2)
Δ df		8	44	3	16
R_1^2 [b]		0.604	0.604	0.606	0.611
R_2^2 [c]		0.664	0.664	0.674	0.680
ICC	0.181				
design effect	3.713				

* *p*<0.1
** *p*<0.05
*** *p*<0.01
[a] centered around grand mean (i.e. mean for entire dataset)
[b] explained variance at level one, the scenario level
[c] explained variance at level two, the person level

I followed the steps described in subsection 4.4.7:

- to test hypothesis 3, regarding the significance and direction of the effects of each of the situation-related factors, i.e. scenario level variables, on the deviation from the normal procedures regarding who should be involved in recovery;
- to answer the research question A3, dealing with the main effect of the person-related factor occupational self-efficacy on the deviation from the normal procedures regarding who should be involved in recovery;
- to answer the research question B3, dealing with the main effect of the person-related factor experience on the deviation from the normal procedures regarding who should be involved in recovery;

- to answer the third part of research question A4, dealing with the possibility of significant interactive effects of the situation-related factors and the person-related factor occupational self-efficacy on the deviation from the normal procedures regarding who should be involved in recovery; and
- to answer the third part of research question B4, dealing with the possibility of significant interactive effects of the situation-related factors and the person-related factor experience on the deviation from the normal procedures regarding who should be involved in recovery.

The corresponding findings are discussed in the remainder of this subsection.

The design effect was again far greater than 2, demonstrating the nested character of the data and the need for a multilevel approach in data analysis. As can be seen from the significant Δ deviance, step 1, which produced model 1, resulted in a significantly better model fit than the empty model, indicating that adding the scenario level variables led to an improved model. Step 2 where the scenario level effects were made random at person level, producing model 2, again resulted in a significantly better model. Also in step 3, where the person level variables main effects were added, producing model 3, Δ deviance was significant, so the model improved significantly. And finally, while some of the cross-level interactions of occupational self-efficacy and experience with the scenario level variables that were added in step 4 to produce model 4 were significant, including the cross-level interactions did not significantly improve the model fit. Therefore, model 3 was used was used to test the hypotheses and answer most of the research questions regarding the deviation from what procedures prescribe regarding who to involve in recovery after detection. Only for answering the research questions dealing with the possibility of significant interactive effects of the situation-related factors and the person-related factors, model 4 will be used, as even though this model did not mean a significant improvement compared to its less detailed predecessor, some significant cross-level interactions were identified.

Similar as in the previous two subsections, for the models presented in this subsection the proportions of explained variance, both at the scenario level and at the person level, were calculated based on formulae provided by Snijders and Bosker (1999, pp. 104-105). The explained variance at the scenario level of the most detailed model, model 4, was 61.1%. The explained variance at the person level of this model was 68.0%. The explained variance at the scenario level of model 3, the last model that meant a significant improvement compared to its less detailed predecessor, was 60.6%. The explained variance at the person level of this model was 67.4%.

As can be seen from the regression coefficients for model 4, for the scenario level variables in the ‘fixed part’ section of table 30, all parts of hypothesis 3 about the factors influencing the deviation from normal procedures regarding who should be involved in further recovery steps after detection were confirmed. The regression coefficients of the all of the eight scenario level variables were all significant into the hypothesized directions. The findings indicated that the greater the normal involvement of others as suggested by procedures or work routines was, the further people who detected a deviation tended to deviate from these routines in the direction of taking on a larger share of the recovery tasks than suggested. Furthermore, situations where those who detected a deviation also had the knowledge and expertise needed to deal with it, where the required tools and equipment were available, and where the necessary recovery steps were more urgent and more important, also shifted this possible procedure deviation in the direction of those who involved in detection taking on a larger than suggested share of the recovery tasks. Situations where much time was needed for the recovery steps, where people were busy with other urgent and important tasks and where others available who could assist were available, led to a shift in the possible procedure deviation in the direction of involving others to a larger extent than suggested. The absence of any of these effects, i.e. in situations where one or more of the scenario level variables would have the value 0, would lead to less forces influencing the possible deviation from the normal routines or procedures.

To enable a comparison of the strength of their effects, the regression coefficients for these eight scenario level variables were standardized, by multiplying the unstandardized coefficients as given in the table above with quotient of the standard deviation of the corresponding variable and the standard deviation of the dependent variable in the model. The effect of what the normal work routines or procedures suggested with regard to who should be involved in the recovery steps was the strongest of all the effects of the scenario level variables (standardized regression coefficient 0.768). In situations where the normal routines or procedures allowed a person who detected a deviation to perform the required recovery steps by him- or herself, procedure deviations could still occur, in the sense that this person would decide to involve others (see negative value of intercept). However, as soon as some involvement of others was suggested by these routines or procedures, such others were never involved to a greater extent than what was suggested. This finding demonstrated that indeed operators were not the kind of people to ‘sit back and wait’ if they felt there was something they could do. The effect of the availability of others who could assist was also quite strong (standardized regression coefficient –0.208). With such others present, the possible deviation would shift a little back towards involving more others. A similar but slightly weaker effect was found for the presence of other urgent and important tasks (standardized regression coefficient –0.141). Apparently these could keep operators from taking on more of the recovery tasks than expected, and lead them to increase the extent to which others were involved instead. Again weaker, but still important, and in the opposite direction, was the effect of having the knowledge and expertise needed for the recovery tasks (standardized regression coefficient 0.076). Armed with such knowledge, operators were ready to take on more than their suggested share of the recovery steps. The effects of the urgency and the importance of the recovery steps were again somewhat weaker (standardized regression coefficients respectively 0.064 and 0.046), with importance being the weakest effect of the two. Nevertheless, they were sufficient reasons for operators to take on more than their suggested share of the recovery steps, if they felt that that was the fastest effective way to intervene. This served as an indication of the (personal) responsibility operators often feel for keeping the process safe. The effect of time required was similar in strength as the effect of importance (standardized regression coefficient –0.044), but in the opposite direction: the possible deviation from the normal routines tended to shift a little back in the direction of more involvement of others. The effect of having the necessary tools and equipment on the procedure deviation was the smallest of all the scenario level effects (standardized regression coefficient 0.041), responsible for a small increase in the direction of those involved in detection taking on a larger than suggested share of the recovery steps.

Given this combination of effects that can lead those involved in detection to take on a larger than suggested share of the recovery steps by themselves, and effects that can lead them to involving others to a larger extent than suggested by the procedures, it is important to take all of the scenario level variables into account to determine which procedure deviation can be expected in a certain situation. Furthermore, such procedure deviations should always be seen in the light of these situation characteristics affecting them. Most of the times, in line with what for example Reason (1990) and Lawton (1998) noted with regard to procedure violations in general, the procedure deviations are in the organization’s best interest, for example when there is no time available to wait for others. It is essential, though, that the people who end up performing the recovery steps know what they are doing. If the people who start the recovery steps do not know how to complete the recovery process, there may at least be a number of things they know how to do, that they can already do, as damage control or to gain time (cf. the descriptions of mitigation and stabilization in subsection 3.3.1).

Research questions A3 and B3 focused on the main effects of the person level variables, respectively occupational self-efficacy and experience, on the deviation from normal procedures regarding who should be involved in the recovery steps after detection. As can be seen from the corresponding regression coefficients, these main effects were significant in model 3, but no longer significant as soon as their cross-level interactions with the scenario level variables were taken into account in model 4. As was explained before, if significant, the main effects of these person level variables could explain the person level variance of the intercept. The commonly accepted practice is to only interpret those main effects in light of the significant cross-level interactions (e.g. Snijders & Bosker, 1999). Even though model 4 with these interactions was not a significant improvement compared to model 3, some of the cross-level interactions were significant. Most likely, the inclusion of all the other, non-significant cross-level interactions affected the overall improvement in model fit. Honoring the significant findings from model 4, I concluded that occupational self-efficacy and experience did not have any main effects. As was explained before, regarding the model building processes for the other dependent variables, an explanation for the lack of main effects of the person level variables included in the follow-up studies could be the homogeneity between the participants regarding these variables. Also in the case of this dependent variable, i.e. the procedure deviation, perhaps the model could have been improved by including other person level variables than those that were actually measured, such as a more task-specific occupational self-efficacy measure, or how people plan their tasks, and which tasks they find most appealing.

As can be seen from the 'random part' section of table 30, all of the scenario level variables had significant variances when made random at person level (from model 2 onwards), which meant that there were differences between the participants with regard to the regression coefficients for these variables. As in the two previous subsections, here, too, an attempt was made to explain these differences by looking for significant cross-level interactions between the scenario-related variables and the person-related variables. The regression coefficients for the cross-level interactions in model 4 show that the only significant cross-level interactions between occupational self-efficacy and the scenario level variables were the interactions with the availability of the required knowledge and expertise, and with the availability of others who could assist. This answered the third part of research question A4, which aimed at identifying interactive effects of scenario level variables with occupational self-efficacy. The regression coefficients for the cross-level interactions in model 4 also show that the only significant cross-level interaction between experience and the scenario level variables was the interaction with the availability of the necessary tools and equipment. This answered the third part of research question B4, which aimed at identifying interactive effects of scenario level variables with experience.

To interpret these interactions, a similar approach was used as in the previous two subsections, that is, two regression equations were compared, one equation with one of the two possible values of the situation-related factor involved filled in in the equation, with another equation in which the other possible value was filled in. The difference between the regression equations for the deviation from what the normal written or unwritten procedures prescribed or suggested with regard to who should be involved in the recovery steps following detection, for a situation where those who detected the deviation had the required knowledge and expertise, and a situation where they did not have the required knowledge and expertise, was 0.230+0.264*ose (based on unstandardized regression coefficients, whereby ose is the score on occupational self-efficacy, centered around the grand mean for the entire dataset). So, this positive difference was larger for people with a higher than average score on occupational self-efficacy, and smaller for people with a lower than average score on occupational self-efficacy, or for them the difference could even become negative. This indicated that based on having the required knowledge and expertise, for people with higher than average scores on occupational self-efficacy , the deviation from normal procedures regarding who should be involved in recovery steps after detection tended to shift further towards them taking on a larger than prescribed or suggested share of the recovery steps, than for their less efficacious coworkers. For people with occupational self-efficacy scores of roughly 0.9 or more under the average score, the direction of the effect of having the required knowledge and expertise on the deviation from what procedures prescribed regarding who should be involved in further recovery even changed, towards them taking on less than their prescribed share of the recovery tasks. Apparently people needed to have at least a basic level of occupational self-efficacy before they were inclined to shift their deviation from normal procedures regarding who should be involved in recovery after detection towards taking on a larger than prescribed share of the recovery tasks, based on having the required knowledge and expertise. Since no main effect was found for occupational self-efficacy, apparently just believing in one's ability to successfully perform the required recovery steps was not enough reason to be inclined to take on a larger than prescribed share of the recovery tasks, this only happened when the people involved also had the required knowledge.

The difference between the regression equations for the deviation from what the normal written or unwritten procedures prescribed or suggested with regard to who should be involved in the recovery steps following detection, for a situation where others who could assist were available, and a situation where no such others were available, was –0.385+0.162*ose (based on unstandardized regression coefficients, whereby ose is the score on occupational self-efficacy, centered around the grand mean for the entire dataset). So, this negative difference was larger for people with a lower than average score on occupational self-efficacy, and smaller for people with a higher than average score on occupational self-efficacy. For the difference to become positive, people would need to have a score of roughly 2.3 higher than the average score on occupational self-efficacy. The dataset did not contain any such cases. These findings indicated that based on the availability of others who could assist, for less efficacious people the deviation from what normal procedures prescribed regarding who to involve in recovery steps after detection tended to shift further in the direction of them taking on less than their prescribed share of the recovery tasks, than for their more efficacious coworkers.

Finally, the difference between the regression equations for the deviation from what the normal written or unwritten procedures prescribed or suggested with regard to who should be involved in the recovery steps following detection, for a situation where those who detected the deviation had the required tools and equipment, and a situation where they did not have the required tools and equipment, was 0.089+0.007*exp (based on unstandardized regression coefficients, whereby exp is the number of years of experience, centered around the grand mean for the entire dataset). So, this positive difference was larger for people with a higher than average number of years of experience, and smaller for people with a lower than average number of years of experience, or for them this difference could even become negative. This indicated that in general, based on the availability of the required tools and equipment, for more experienced people the deviation from what the normal procedures prescribed regarding who should be involved in the recovery steps after detection tended to shift further in the direction of them taking on more than their prescribed share of the recovery tasks, than for their less experienced coworkers. For people with roughly 13 or more years less experience than average, the direction of the effect of having the required tools and equipment on the deviation from what procedures prescribed regarding who should be involved in further recovery even changed, towards them taking on less than their prescribed share of the recovery tasks. Apparently, at least a basic level of experience was necessary for people to be inclined to shift their level of involvement in further recovery after detecting a deviation towards taking on a larger than prescribed share of the recovery tasks, based on having the necessary tools and equipment available.

4.5.6 Summary

In line with what was known from time management research (e.g. Koch & Kleinmann, 2002; Claessens et al., 2003) regarding tasks in general, in the two studies presented here, urgency and importance were found to be the most important factors in task prioritization for recovery tasks, whereby urgency had a slightly stronger effect. Many of the expectations stemming from action theory (cf. Heckhausen & Kuhl, 1985) regarding how likely it is for intentions to be transformed into actions were confirmed with specific regard to recovery tasks: ordered from strongest to weakest effect, the presence of other tasks requiring immediate attention, lacking the necessary tools and equipment, and the absence of others who could assist were all found to delay the start of the required recovery steps. As predicted, lower priority was also given to recovery tasks where the suggested role of others than those who detected the deviation was larger. No time discounting effect (cf.

Koch & Kleinmann, 2002) was found, apparently people did not let the thought that recovery actions would take a long time to complete deter them from getting these actions started. No effect was found either of the availability of the knowledge and expertise required to perform the recovery tasks on the priority given to these tasks, people apparently knew which priority these tasks deserved even without knowing how to perform them. Furthermore, while there were no main effects of occupational self-efficacy and experience, the findings indicated that less efficacious people may not have as much confidence as more efficacious coworkers to delay recovery actions in favor of other urgent and important tasks; that more experienced people may not allow other tasks to delay them as long as less experienced people; and that less experienced people depended more on the assistance of others to get the recovery steps started.

With regard to how the recovery tasks that were added to the already existing workload were distributed, this was indeed to a large extent dictated by what the procedures or unwritten work routines prescribed (cf. subsection 3.3.5). Furthermore, in line with what was noted by Waller, and findings presented in subsection 3.3.5, the research showed that people tended to involve more others in cases where at least others who could assist were available, where other tasks also required immediate attention, where they did not have the required knowledge and expertise, where recovery was less urgent and less important, where they lacked the necessary tools and equipment, and where greater amounts of time were needed for the recovery tasks. The influence of especially the first three of these factors was substantial. In addition, experience had no main effect on recovery task distribution and the positive main effect of occupational self-efficacy had to be viewed in the light of the interactions of this factor with situation-related factors. The findings indicated that for more efficacious people, having the required knowledge and expertise was a stronger reason to involve fewer others in recovery, than for their less efficacious coworkers. For less efficacious people, the availability of others was a stronger reason to involve more others in recovery, than for their more efficacious coworkers: less efficacious people would more readily use such others whereas more efficacious people may have felt the responsibility to stay more involved. For less experienced people, having the required knowledge and expertise was a stronger reason to involve less others, than for their more experienced coworkers: possibly less experienced persons felt a greater need to practice, as opposed to a need to share their experience. For more experienced people, having the required tools and equipment was a stronger reason to involve less others, than for less experienced coworkers: possibly as their experienced increased, people felt they knew better how to handle such tools without much assistance from others. Also, a basic level of experience was needed before people were inclined to allow the availability of the required tools and equipment to shift the extent to which they would involve others in further recovery towards lower levels.

As for the extent to which people who detected a deviation would deviate from what the procedures or unwritten work routines prescribed or suggested regarding who should perform the recovery tasks: the greater the prescribed involvement of others, the further people who detected a deviation tended to deviate from this prescribed involvement, by taking on a larger than prescribed share of the recovery tasks themselves. People were also more likely to take on more than their prescribed share of the recovery tasks in situations where no others were available, where they were not busy with other urgent and important tasks, and where they had the knowledge and expertise required for the recovery tasks. To a lesser extent, this tendency also increased in situations where the necessary recovery steps were more urgent and more important, where the required tools and equipment were available, and where little time was required for the recovery steps. Opposite values of these factors each tended to shift the possible procedure deviation in the direction of giving a larger than prescribed share of the recovery tasks to others. Furthermore, while there were no main effects of occupational self-efficacy and experience, the research indicated that the more efficacious people were, the more having the required knowledge and expertise increased their inclination to take on a larger than prescribed or suggested share of the recovery tasks. Apparently just believing in one's ability to successfully perform the required recovery steps was not enough reason to be inclined to take on a larger than prescribed share of the recovery tasks, this only happened when the people involved also had the required knowledge. The research also indicated that that less efficacious people were more inclined to involve others to a larger extent than prescribed when they were available; and that only with enough experience people would be inclined to use the tools and equipment that should officially be used by others for the recovery tasks.

4.5.7 Limitations

As can be seen from the findings discussed above, the follow-up studies were successful in determining the separate effects of each of the included scenario level variables on task prioritization and distribution in situations where recovery is needed, while controlling for the other effects. The person level variables that were included in the studies, however, were much less successful in explaining the variance between the different participants. Only on the dependent variable of extent to which others were involved in recovery, did one of the person level variables, occupational self-efficacy, have a significant main effect. For each of the three dependent variables, there were a few significant cross-level interactions. The setting in which the studies were performed may have limited the amount of significant findings involving the person level variables. The participants formed a very homogeneous group with regard to both occupational self-efficacy and number of years of relevant work experience. A possible way to overcome the problems associated with this homogeneity would be to also include participants from different organizations where such homogeneity is not expected. To find a better explanation for the interpersonal differences found within the studies described in this chapter, perhaps the models for the dependent variables could have been improved by including other person level variables than those that were actually measured. A more task-specific occupational self-efficacy measure (cf. Eden & Kinnar, 1991), for example, may result in larger differences between participants. The participants may also differ in their preferences in task planning, and regarding which tasks they find most appealing, two factors that can also be expected to influence their decisions regarding task prioritization and distribution.

Furthermore, the choice for a specific type of chemical process plant where the follow-up studies were performed may have limited the generalizability of the findings to some extent. But as was demonstrated in the discussion of the findings, it was possible to develop models for the dependent variables that applied in both of the participating plants, so the findings were at least proven not to be organization-specific. If the studies were to be repeated in different types of environments, the abstract descriptions of the characteristics of the vignettes would not have to change, but some of the examples might need to be changed as some of them were rather production process-specific. The participants were instructed to base their answers on the abstract descriptions, so the answers would apply to any situation with these characteristics and not just the example provided. Provided they acted according to this instruction, or at least to a reasonable extent, a comparison would be possible of the findings obtained so far with findings from possible additional studies. While I expect the same variables to play a role in task prioritization and distribution, the weights of the regression coefficients may differ between different work environments.

Then, there is the issue whether or not the particular subset of vignettes given to a participant, and the order in which vignettes and factors within these vignettes were presented to the participants, may have affected their responses. Since the subsets for the participants were combined via a random selection process, this cancelled out any possible effects of the order in which the vignettes were presented on the dataset as a whole. This random combination process also cancelled out any potential effects of the combination of vignettes presented to a participant. The order in which the situation-related factors within a vignette were presented, however, was not random, but the same for all vignettes. If there had been any effects of the position in the vignette where a factor was presented, this would probably have been visible in the regression coefficients for the corresponding factors. Stronger effects on the dependent variables would have been found either for factors at the top of the description, or otherwise for factors at the bottom, which are closest to the questions, so possibly easiest to remember. The regression coefficients revealed no such tendencies. In fact, for the priority given to recovery, the factors in the middle of the description had the strongest effects. Furthermore, for none of the dependent variables, rank ordering the factors according to the strengths of their effects resulted in the same following order as the description and neither in an exactly opposite order. Therefore, it is not likely that any strong order effects were present in the data.

Finally, what the participants in the studies indicated that they would do given a certain scenario might have been different from what they would actually do when confronted with such a situation in practice. Their answers might have been more in accordance with procedures than their actual behavior would be, thus limiting the opportunities to identify circumstances where they would deviate from what normal procedures prescribe with regard to who should be involved in further recovery after detection (i.e. the third dependent variable). This possibility of differences between what people say and do, of getting socially desirable responses, has also been recognized by, for example, Rossi and Anderson (1982) and by Liker (1982). By comparing judgments of real-world situations with judgments of situations described in vignettes, the latter found that vignette analysis did not bias the mechanisms underlying judgment processes. To minimize the possibility of getting answers that do not correspond what people would to in practice, in the introduction sessions (see section 4.4) every effort was made to ensure that participants would answer the questions they were asked truthfully, by emphasizing the anonymity of their responses. These efforts must have paid of to some extent, as the analysis results showed that there were indeed circumstances possible under which people would deviate from procedures regarding who should be involved in recovery. Next to the possibility of people giving a socially desirable answer on purpose, even people who did not mean to give an answer that did not correspond with what they would actually do, might have had difficulties in predicting their response without actually being in the situation described. The possibility of getting responses that did not correspond with what people would actually do, however, was even further reduced by including only scenarios representing credible situations in their work environment, which almost all of the participants were likely to have experienced before, either first- or second hand, given their many years of experience in operations.

4.5.8 Future research directions

The discussion above points to a number of areas that would benefit from further research. First, it seems worthwhile to look for other person level variables that may do a better job than those that were used so far to explain the identified interpersonal differences in task prioritization and distribution. Insight in such person level variables may be useful for training and selection of personnel. Another area worth exploring is what the effects are of the same variables as those used so far on task prioritization and distribution in other organizations. That way, the generalizability of the findings can be established, or differences between different types of work environment can be identified. A possible additional benefit of including other types of organizations in the research is that this may produce more variance with regard to the person level variables, increasing the chance that these can explain interpersonal differences. Finally, to further reduce the possibility of getting responses that do not correspond with what people would actually do in a certain scenario, and to get more information about the entire recovery process that would follow, researchers could ask participants, either on an individual basis or as teams (cf. Crew Resource Management (CRM) training with airline crews, see for example Helmreich, Merrit, & Wilhelm, 1999) to talk them through all the steps they would take to recover from a situation presented to them in a written scenario. What is more, this would create the possibility for the researchers to add additional influences at later recovery process stages, and to ask for further information and clarification where needed.

The implications of both the follow-up studies presented in this chapter and the initial studies presented in chapter 3 will be summarized in the next and final chapter. A detailed discussion of what these implications mean with regard to future research can be found there, too.

5 General discussion and conclusions

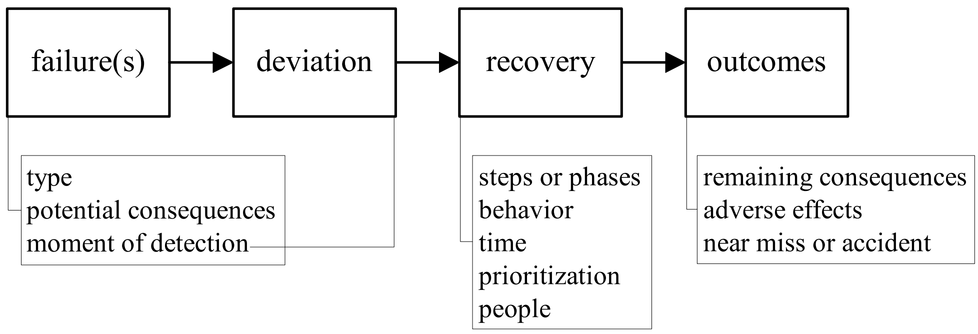
5.1 Purpose and outline

The lessons learnt from the research described in this thesis encompass different areas. Among them, not only theoretical, but also methodological, and practical implications can be distinguished. Furthermore, related to these implications, a number of suggestions can be made for future research. In this chapter, I will address these implications and suggestions, in the same order in which they were mentioned above.

5.2 Implications of the research described in this thesis

5.2.1 Theoretical implications

Recall from chapter 1 that the main objective of the research described in this thesis was to gain insight into the processes people follow to recover from different types and combinations of failures in work situations, and the factors that influence these processes and their outcomes. To guide the research, I made and used the following model of the sequence from failures, leading to a deviation, which is detected at some point and further recovered, eventually resulting in certain outcomes, that depend on how successful the recovery process has been:



The more detailed research questions that were presented in chapter 2 of this thesis aimed at gaining further insights into specific aspects of the main elements of this sequence, and into how these aspects are interrelated. The first five specific research questions covered the recovery process itself, as main research topic, they aimed at establishing what is done, why, how, when, for how long, by whom, and why by those persons, during a recovery process. The sixth specific research question aimed at explaining what happens during a recovery process (cf. the first five research questions) by establishing how the preceding types and combination of failures, the potential consequences of the deviation to which these failures led, and the moment of detection influenced the recovery processes that follows. The seventh and last specific research question aimed at establishing the recovery process end results and how these can be explained by both the events leading up to the recovery process (cf. research question six) and the recovery process itself (cf. the first five research questions).

The first specific research question was: *a) With regard to the steps and transitions involved, does the proposed recovery process phase model accurately describe the processes people follow to recover from failures in work situations, b) which different types of recovery processes can be distinguished based on which phases occur when during the process, and c) which additional details, in the form of subgoals, can be distinguished for each of the phases of the proposed phase model?* With regard to the existence of three types of phases or steps involved in recovery, namely detection, explanation, and countermeasures, in the initial three studies, described in chapter 3 of this thesis, the proposed recovery process phase model was verified empirically. Furthermore, while existing insights (e.g. Zapf & Reason, 1994; Johannsen, 1988) already demonstrated the existence of these three phases, and some researchers (e.g. Van der Schaaf, 1988; Bagnara et al, 1988; Frese, 1991; Kontogiannis,1999) had pointed out that the explanation phase does not necessarily have to occur, the initial three studies added to these insights by confirming that the transitions between and repetitions of certain phases that were proposed in the process phase model were indeed possible. A recovery process always started with *detection*, and after that, any number and combination of steps aimed at *explanation* and *countermeasures* were possible, ranging from no such steps occurring at all, to processes involving multiple steps for each of these two phases, including iterations. Since a lot of recovery processes did not involve any efforts spent on explanation, and those that did varied with regard to the moment in the process when such efforts were made, the occurrence and timing of explanation phases were used to distinguish between different types of processes. In addition, based on the specific goals of separate recovery steps, further detail was added to the model, similar to how this was done by, for example, Norman (1984), who distinguished separate stages and levels of activity in human-machine interaction, and Newell and Simon (1972), who distinguished separate episodes in problem solving, each with their own subgoal. In the explanation phase two sub-phases were distinguished, definition of the problem, and identification of the causes, and within the countermeasures phase four sub-phases, stabilization, mitigation, temporary correction, and permanent correction. Again, not all of these sub-phases needed to occur within any given recovery process, and iterations of sub-phases were possible too. The newly identified sub-phases proved to be useful in establishing the reasons why certain recovery steps were performed at certain stages in the process.

The second specific research question was: *For each of the distinguished types of recovery process phases (i.e. detection, explanation, and countermeasures), a) which proportion of all the steps belonging to that phase involves ad-hoc behavior and which proportion pre-planned behavior, and, b) furthermore, which proportion involves skill-based behavior, which proportion rule-based behavior, and which proportion knowledge-based behavior?* The distinction between skill-based, rule-based, and knowledge-based behavior has been used widely in the study of errors (e.g. Reason, 1990), and, via the initial three studies described in this thesis, was shown to apply to recovery process steps as well. The difficulty in coding for this behavior was that one recovery step could involve performance at several behavior levels, for example steps where a specific goal and a plan for reaching it were established at knowledge-based level, but once this was done many of corresponding actions could be performed at rule-based level. In such cases, the entire process step was coded as involving the highest of the applicable behavior levels. As every behavior level comes with specific demands as to how to optimally support such behavior, the analysis of the behavior levels involved in the recovery steps provided specific leads for optimizing recovery possibilities within an organization. In keeping with the large share of recovery steps involving knowledge-based behavior found in the studies, the research highlighted the important contribution to recovery steps of ad-hoc behavior, not specifically pre-planned by the organization through formal or informal procedures and work routines or through training. While especially for those countermeasures where fast reactions were essential, pre-planning these recovery steps was found to be useful and to occur frequently; on the other hand, especially for the explanation phases but also to detect deviations, operators were often found to use their detailed process knowledge in a more ad-hoc manner.

The third specific research question was: *How long do a) different types of recovery steps at various stages in the process, and b) the entire process take?* In the initial three studies, the earlier stages in the process were shown to be completed the fastest, as they were often meant to get things more or less back under control. The overall amount of time that passed between process start and completion was shown to be longer than merely the sum of the amounts of time spent on each of the separate recovery steps, as a result of time elapsed in between those steps, spent either waiting before the next step was possible, or working on other tasks, or both.

The analysis of how tasks are prioritized, and distributed across people, was relatively new in recovery-related research. The fourth and fifth specific research questions dealt with these issues. These questions were, respectively: *Based on which factors do people* *prioritize tasks when a deviation needing recovery presents itself at work?* and: *If others than those already involved become involved in a recovery process at work: a) what are the organizational roles of these people, b) when do they become involved, c) for what reasons do they become involved, and d) what is the effect of their involvement? Furthermore, if those involved in detection do not involve any others in the rest of the recovery process at all, for what reasons does this happen?* The three initial studies presented in this thesis showed that the importance of newly added recovery tasks, operationalized in a measure combining type and severity of their potential consequences, and the urgency and importance of other already ongoing tasks, influenced the priority given to further recovery steps after the detection of a deviation. The initial studies also confirmed other researchers' findings (Doireau et al., 1997; Hutchins, 1996; Wioland & Amalberti, 1998) with regard to the important role in the detection phase of others than those who were involved in the failures, and further demonstrated that the involvement of others most often remained essential in the rest of the recovery process as well, because of specific qualifications, skills and expertise these other people had, or the additional resources they brought to the scene. If no others than those detecting the deviation were involved in the subsequent recovery process, this was because the person who completed the recovery process alone was allowed to perform all the necessary steps, and responsible for performing the necessary steps, and, even if the other two conditions were not applicable, at least also able to perform the necessary steps alone.

From existing literature (e.g. Waller, 1999, 1997) it was clear that task prioritization and distribution were important responses of people when confronted with unexpected events, or, more specifically, with deviations. Furthermore, time management research (e.g. Covey et al., 1994, Koch & Kleinmann, 2002) and action theory (e.g. Heckhausen & Kuhl, 1985), for example, had provided insight into factors that influence task prioritization and distribution. The two follow-up studies presented in this thesis combined these insights with the findings from the initial studies, to develop more detailed models of task prioritization and distribution specifically in situations where recovery from a deviation was needed.

Urgency and importance of the recovery tasks were shown to be important factors in prioritization in such situations, which was in line with what, for example, Koch & Kleinmann (2002) argued for task prioritization in general. In addition, the presence of other tasks requiring immediate attention, lacking the necessary tools and equipment, and the absence of others who could assist, were all found to delay the start of the required recovery steps, but their effects were less strong. Lower priority was also given to recovery tasks where the suggested role of others than those who detected the deviation was larger. No time discounting effect (cf. Koch & Kleinmann, 2002) was found, apparently people did not let the thought that recovery actions would take a long time to complete deter them from getting these actions started. No effect was found either of the availability of the knowledge and expertise required to perform the recovery tasks on the priority given to these tasks, people apparently knew which priority these tasks deserved even without knowing how to perform them. Furthermore, while there were no main effects of occupational self-efficacy and experience, the findings indicated that less efficacious people may not have as much confidence as more efficacious coworkers to delay recovery actions in favor of other urgent and important tasks; that more experienced people may not allow other tasks to delay them as long as less experienced people; and that less experienced people depended more on the assistance of others to get the recovery steps started.

The distribution of recovery tasks was to a large extent dictated by what procedures prescribed in this regard. Furthermore, the research showed that people tended to involve more others in cases where at least others who could assist were available, where other tasks also required immediate attention, where they did not have the required knowledge and expertise, and in cases where recovery was less urgent and less important, where they lacked the necessary tools and equipment, and where greater amounts of time were needed for the recovery tasks, even though the last four effects were somewhat smaller. Occupational self-efficacy reduced both the effect of the availability of the required knowledge and the effect of the availability of others on the extent to which people involved others in recovery. Experience increased the effect of the availability of the required knowledge and decreased the effect of the availability of the required tools and equipment on the extent to which people involved others in recovery.

The research demonstrated that in certain circumstances people would deviate from what normal procedures or work routines prescribed regarding who should be involved in the recovery tasks. The greater the prescribed involvement of others, the further people who detected a deviation tended to deviate from this prescribed involvement, by taking on a larger than prescribed share of the recovery tasks themselves. People were more likely to take on more than their prescribed share of the recovery tasks in situations where no others were available, where they were not busy with other urgent and important tasks, and where they had the knowledge and expertise required for the recovery tasks. To a lesser extent, this tendency also increased in situations where the necessary recovery steps were more urgent and more important, where the required tools and equipment were available, and where little time was required for the recovery steps. Opposite values of these factors each tended to shift the possible procedure deviation in the direction of giving a larger than prescribed share of the recovery tasks to others. These procedure deviations were not necessarily a bad thing, as the intention of those involved was to get the job done (cf. Lawton, 1998) and the intial studies showed that the result often was that negative consequences were avoided. Of course if people deviate from what procedures prescribe or normal work routines suggest, it is vital that those who end up performing tasks that are not theirs know what they are doing, otherwise additional risks arise, associated with the violation of procedures (cf. Reason, 1990; Lawton, 1998). In this regard, the research indicated that just believing in one's ability to successfully perform the required recovery steps was not enough reason to be inclined to take on a larger than prescribed share of the recovery tasks, this only happened when the people involved also had the required knowledge. The research also indicated that that less efficacious people were more inclined to involve others to a larger extent than prescribed when they were available; and that only with enough experience people would be inclined to use the tools and equipment that should officially be used by others for the recovery tasks.

In the sixth specific research question, the scope of the research was widened to include the events preceding a recovery process, to establish their effects on this process: *How are a) the failure types that precede a recovery process in a working environment, b) their potential consequences, c) the moment of detection (how long the deviation already existed, and performance stage when errors are detected), and d) the form the recovery process that follows takes with regard to steps, behavior, and people involved, prioritization and time spent, interrelated?* And finally, the seventh and last specific research question aimed at bringing together all the aspects of the recovery process and the preceding events discussed so far. In this question, the scope of the research was extended once again, this time to also include the situation after the recovery process is completed, the process outcomes or end results, to establish how these were influenced by everything that preceded them: *What is the relationship of a) the failure types that precede a recovery process in a working environment, b) their potential consequences, c) the moment of detection (how long the deviation already existed, and performance stage when errors are detected), and d) the form the recovery process that follows takes with regard to steps, behavior, and people involved, prioritization and time spent, with recovery process outcomes?*

The analysis of how failures, various aspects of the recovery processes that followed, and the process outcomes or results were interrelated, asked for by these last two research questions, was fairly new in the domain of safety and reliability, too. Existing research had only attempted to do this for errors, i.e. human failures, made at operational levels, not for other types of failures or combinations thereof, and only for some, but not all of the aspects of recovery processes distinguished in the studies presented here. No support was found for findings from, for example, Bagnara et al. (1988), and Sellen (1994), that skill-based errors are easier to detect than rule- and knowledge-based errors, possibly as a result of the small number of cases involving only human failures. In line these other researchers' findings, recovery from knowledge-based errors was found to involve more efforts aimed at explanation than recovery from skill- or rule-based errors.

Stepping outside this commonly found restriction to recovery from human failures, a few typical patterns emerged from the data from the three initial studies: First, the presence of technical failures, i.e. failures of or defects in technical systems, installations, equipment, tools, and materials, and organizational failures, made at management levels in the organization, increased the need for explanation. The presence of organizational failures also delayed the moment of detection. Furthermore, late detection led to longer processes in which more people were involved, such processes also involved more steps, and as result of all of these things combined, more, or more severe consequences remained at the end. And finally, cases involving technical and/or organizational failures and/or more, or more severe potential consequences led to longer processes involving more steps and people, than cases involving only human failures and/or few or minor potential consequences, and as a result of all these things combined, more, or more severe consequences remained after the recovery process.

5.2.2 Methodological implications

Various data collection and analysis methods were used in the studies presented in this thesis, each with their own advantages and disadvantages. Now that the studies have been completed and been presented in the previous chapters, this is a good moment for a reflection on a few methodological issues.

The use of data from an existing near miss reporting system, as in the first study, proved to be very helpful for collecting a relatively large amount of data within a relatively short time. A word of warning may be in order, as my experience was that reports from such a system do not necessarily contain much information about the recovery process that was followed and that ensured that the event turned out to be a near miss instead of an accident involving negative consequences. To get such information I had to do follow-up interviews with those involved in the events. Nevertheless, the reports still formed a good starting point and at least covered some of the information I was looking for, most notably with regard to what went wrong, i.e. the failure process that triggered the recovery steps. Furthermore, I found that people in organizations with a near miss reporting system in place tended to have quite a firm grasp of what near misses are and how recovery steps play a role in these, which saved a lot of effort by not having to explain all of this to them first. The organization that participated in the third study did not have a reporting system specifically aimed at collecting data about near misses, and I found that I had to put a lot more effort in explaining what kind of information I was after, there. This difference between the two organizations, that is, having successfully established a near miss reporting system or not, may have impacted the thresholds of what was reported in study 1 and what in study 3 as well. The cases collected in study 3 involved more, or more severe potential consequences than those in study 1.

Event diaries, used in the second study, proved to be helpful in getting information about very specific types of events. Having completed the first study helped a lot in deciding which data to ask for in the diary proformas. The main advantage of diaries over doing interviews, or follow-up interviews triggered by regular near miss reports, is that it takes far less effort from the researcher to collect the data (cf. Breakwell & Wood, 1995). The main disadvantage, however, is that the success of the study depends much more on the willingness of the participants to contribute (cf. Symon, 1998). Most of the participants in my studies found talking to me for a few minutes less of a hassle than filling out a proforma. I encouraged the operators to participate by making the objectives well understood, and keeping in regular contact. These are two of the ways to increase participation mentioned by Symon (1998), next to these she also mentioned making the diaries personally useful. Furthermore, I found that while in interviews lack of clarities in a person's event description can be addressed instantly, lack of clarities in events reported via a diary may require additional site visits to talk to the persons who documented those events. Finally, the threshold of what was reported appeared to be lower in study 2 than in both other studies: the cases collected in this study involved less, or less severe potential consequences. To some extent, this may have been the result of the fact that in this study, information was asked about specific types of cases that were normally not reported.

The data coding method developed for and used in the initial three studies proved to enable comparison of a wide variety of recovery processes with regard to a wide variety of aspects. The more than satisfactory scores of the coding method in interrater reliability tests inspire confidence in the method's objectivity and usability by others and in other environments.

In the initial three studies, all the recovery processes about which data was collected were 'one of a kind', that is, even though a lot of very detailed information was obtained about every single one of them, it was unknown if others when confronted with the same situation would have responded in the same way. In the follow-up studies, the vignette approach overcame this problem, but at the expense of a lot of detail in the participants' responses. Given the type of information I was looking for in studies 4 and 5, this was not a problem, rather an intentional result, to facilitate data analysis, given a few very specific research questions. In fact, similar vignettes as the ones used in studies 4 and 5 could be used to collect much more detailed information as well, about what people would do, or at least say they would do, when confronted with the described situation. To this end, a researcher could present a person or team of coworkers with a specific vignette, and ask them to talk him or her through all of the recovery steps they would take to deal with such a situation. The vignettes could be distributed over participants in a similar way as was done in the follow-up studies.

Of course the quality of the responses to vignettes depends on how well the participants succeed in imagining themselves in the situations described in the vignettes (cf. Rossi & Anderson, 1982). Even if nothing keeps them from answering truthfully, such as worries how their answers will be used, which may lead to socially desirable responses, there is always a chance that when actually facing the described situation, they would do something different than what they said they would. By involving local experts in the development of examples corresponding with the different scenarios, I made the described situations as realistic as possible, thus minimizing the chances that people could not relate to the vignettes. For an even more realistic representation of a certain scenario, the use of simulators (e.g. Grau, Doireau, & Poisson, 1998) can be beneficial, but these are expensive and the development of complete scenarios to be acted out in them is very time-consuming.

While the multilevel approach in the analysis of the data from the vignette studies at least did justice to the obvious and demonstrated nested character of the data, i.e. scenarios nested in participants, the person level variables that were included in the studies were not very successful in explaining the variance that was found between participants. More research is needed to identify person level variables that are better at explaining this variance. Insight in such variables may be useful for training and selection of personnel.

In general, the combination of the different data collection and analysis methods used in the research described in this thesis was one of its strong points. After all, if different methods using different sources lead to the same findings, this gives those findings more credibility, and ensures construct validity (cf. Yin, 1994). Furthermore, repetition of the same approach in different studies and the use of a second rater in the coding process in the initial studies helped to ensure reliability of the findings. Finally, the replication principle (Yin, 1994) that was followed to select organizations to participate in the research allowed for further generalization of the findings and thus helped to increase their external validity.

5.2.3 Practical implications

Next to the scientific implications described so far, the research presented in this thesis had some practical implications as well.

The initial three studies, for example, provided insight into what organizations can do to *learn from recovery*. If their near miss- and more serious incident reports were to include information on the recovery process followed, that is, as long as some recovery steps were taken, such information could be used for sharing knowledge within the organization or industry on what to do if similar events occur. The near miss and incident reports could thus contribute to organizational learning (e.g. Argyris, 1999; Huber, 1996) in more than one way. The traditional ways in which incident reporting systems have mostly been used so far for organizational learning were to trigger corrective actions, and to provide a basis for measures to prevent recurrence of the incident (e.g. Van der Schaaf, Lucas, & Hale, 1991; Koornneef, 2000; Hale, Wilpert & Freitag, 1997; Turner & Pidgeon, 1997, chapter 11). By including information about the recovery steps taken, the factors that influenced these steps, and their effects, organizations and their employees can learn about successful recovery strategies and about situations for which better strategies have to be found, too. The recovery process phase model and the data collection and coding methods presented in this thesis can be used for guidance regarding which data to collect.

Focusing on recovery steps in near miss- and more serious incident reports also means the inclusion of much more positive aspects of the incident than the factors that caused things to go wrong. Such a positive point of view during the collection of information about and the analysis of incidents may even help to *overcome barriers to reporting* (see Van der Schaaf & Kanse, 2004, for an overview of such barriers).

The studies described in this thesis also provided some indications regarding what organizations can do to *improve their recovery possibilities*. Of course, as was mentioned above, capturing information about recovery processes that take place in the organization is one way to ensure that knowledge about these processes is spread throughout the organization, so that everyone becomes aware of available recovery possibilities. Furthermore, the research showed that knowledge-based and unplanned or ad-hoc behavior of the people involved plays a very important role in recovery processes, especially in detection phases and explanation phases. This indicated the importance for organizations to invest in their employees' detailed process and plant knowledge to prepare them even for the unexpected. Such knowledge helps them to detect and analyze deviations, and is also helpful for developing custom-made or tailoring known countermeasures, and thus improves recovery possibilities. For possible deviations where fast recovery is essential, organizations should provide pre-planned countermeasures, which were shown to speed up the process. The total amount of time elapsed between recovery process start and completion was shown to not only depend on the time spent on each of the recovery steps, but also on the time elapsed in between those steps. Therefore organizations should at least strive to get the potential hazards associated with a deviation contained within as little time as possible, early on in the process, before other tasks are allowed to delay the progress of the recovery process. To ensure that people make the correct decisions in task prioritization, organizations should provide them with the knowledge required to make correct assessments of the urgency and importance of both recovery and normal tasks. In addition, since the role of others, not only for the detection of a deviation, but throughout the entire recovery process, was shown to be important in many recovery processes, organizations should ensure that access to and communication with such others is made as easy as possible. Depending on an organization's current minimally required manning levels, an evaluation of whether these levels need to be adjusted, in light of what is required for effective responses to any unwanted deviations, might be in order.

Furthermore, since late detection was shown to lead to recovery processes that take more time and the efforts of more people, and end with more, or more severe remaining consequences, organizations may want to focus on reducing this time as much as possible. Organizational failures, made at management levels in the organization, were often only detected after a longer period of time. Compared to cases involving only human failures (made at operational levels) and few or minor potential consequences, cases involving technical and/or organizational failures and/or many or severe potential consequences were shown to lead to longer processes, involving more steps and people, and more, or more severe consequences remain afterwards. This indicated that there may be some room for improvement regarding the ability to recover from such failure combinations. This does not mean that recovery from only human failures does not deserve any additional attention, since the potential consequences of such failures if no recovery steps are taken can also be very serious and organizations should be careful not to lose their expertise in this area.

To facilitate recovery, organizations can use available information- and communication technology to support separate recovery steps or entire processes (e.g. Kanse & Van der Schaaf, 2000c). This possibility has been pointed out, too, by the Institute of Medicine (in press), who also underlined the importance of improving recovery possibilities and learning from recovery for the health care domain. Communication technology offers, for example, a wide range of means to contact others when their assistance is needed. Information technology offers the chance to provide computer support for a wide variety of tasks. It can help in monitoring processes to detect deviations – as long as warnings with regard to alarm overload (e.g. Elzer, Weisang, & Zinser, 1993; Stanton & Baber, 1995) are taken into account and ergonomic design principles (e.g. Salvendy, 1997; Wickens et al., 1998; Kroemer, Kroemer, & Kroemer-Elbert, 2001) are followed. Information technology, in the form of performance support -, decision support -, or expert systems (e.g. McGraw, 1994; Carr, 1992; Wickens et al., 1998), can also help operators to search for likely causes and available solutions. Information technology can even be used to reduce the task load of operators when needed (e.g. Millot, Taborin, Kamoun, & Willaeys, 1988; Rouse, 1988; Parasuraman, Molloy, & Singh, 1993). When such support is implemented, however, cautions from, for example, Bainbridge (1987), regarding the ironies or unintended side-effects of automation, should be kept in mind.

Another way to improve an organization's chances in recovery is via training. With regard to recovery from technical failures, many relevant insights are already available from studies that focused on fault diagnosis training (e.g. Patrick & Stammers, 1981; Duncan, 1981, 1987a; Kostopoulou & Duncan, 2001). Information technology offers the possibility to simulate work situations without affecting the ongoing, real production processes. In simulators (e.g. Grau et al., 1998; Shepherd, 1986; Mehl & Schuette, 1999; Farmer, Van Rooij, Riemersma, Jorna, & Moraal, 1999), employees can practice their responses to deviations that do not occur regularly. A less costly alternative would be to use written scenarios describing such deviations, like the ones used in studies 4 and 5, for training purposes, whereby people cannot always act out their responses, if this would interfere in the production process, but at least the whole response can be discussed in detail.

When organizations start recording information about recovery processes, they may also want to monitor their performance in this area. To this end, they have to establish how they want to *measure the success of a recovery process*. The recovery process end result or outcomes may not be the only thing that matters, organizations may also want to take the potential consequences that were avoided into account, and the resources spent on the recovery process, and possibly some additional aspects that may be particularly relevant to a specific organization as well. If recovery performance information is available, organizations can establish in which areas this performance is lacking, decide on and implement actions to improve this, and afterwards establish what effect these efforts have had.

5.3 Future research directions

To add to the newly gained insights into recovery processes, a larger dataset of recovery processes in combination with the appropriate statistical techniques could possibly reveal more details concerning which of the failures in cases involving a combination of preceding failures had the most influence on the way the recovery process turned out. Furthermore, by including more cases involving unsuccessful recovery attempts, the analysis could show which factors determine the difference between successful and failed recovery.

Future research could also include other organizations in new empirical studies, to enable further generalization of the research findings to other types of organizations, not just chemical plants. Particularly interesting could be the inclusion of non-high reliability organizations, which would allow comparison of recovery processes between the different types of organizations.

When measuring the success or effectiveness of recovery processes, decisions need to be made with regard to the factors that are important to consider in this measure. In my initial three studies, to operationalize effectiveness I mainly used the remaining consequences, and furthermore the adverse effects of well-intended recovery steps, if any. However, organizations may also find it important to look at the man-hours and financial resources spent on recovery, and may want to take into account what was avoided through the implementation of the recovery steps. Following performance measurement principles (e.g. Kaplan & Norton, 1996; Pritchard, 1990), based on all of these factors it should be possible to determine how successful a recovery process has been. In addition to measuring the overall success of the entire recovery process, it may also be useful to determine how successful separate recovery steps have been, to what extent their goals were achieved. To visualize the effects of every separate recovery step, it may be useful to plot a number of graphs displaying what happens over time as the separate steps are taken, for example with regard to potential consequences (should decrease as recovery process progresses), or with regard to the time available until the earliest possible occurrence (should increase), or with regard to the likelihood of negative consequences actually occurring (should decrease).

Another possibility worth exploring involves an alternative use of figure 3 (see section 2.1, or the beginning of this chapter), which illustrated the sequence from failures to a deviation, subsequent recovery, and end results, and which has been used throughout this thesis as a point of reference with regard to which issue was being explored in each section. This figure could also be used as a basis for models in which the relationship between two variables is expected to be influenced by a third variable. In such models, several aspects of the recovery process can be seen as third variables, influencing the effect of the preceding failures (independent variable) on the outcomes or end results of a recovery process (dependent variable). If they actually have an effect, these third variables can either have a moderator function or a mediator function. According to definitions given by Baron and Kenny (1986), a variable with a moderator function partitions a focal independent variable into subgroups that establish its domains of maximal effectiveness in regard to a given dependent variable, and a variable with a mediator function represents the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest. For the analysis of those moderator or mediator effects, a comprehensive success measure for recovery process end results, as suggested in the previous paragraph, would be very useful, too. Within the scope of this thesis, no attempts have been made to test whether models with moderator or mediator variables would fit the collected data. Only via combining relationships into larger patterns based on sequence in time, inferences were made regarding interrelations between three or more variables. While models including moderator or mediator variables could be very useful in identifying interrelationships between multiple variables, they may require more data than the relatively small number of cases collected so far in the initial three studies. In addition, they would require the use of sophisticated statistical data analysis methods to deal with the categorical or at best ordinal nature of the variables that were measured regarding the recovery process and the preceding failures.

With regard to the vignette studies, future research could focus on locating factors at the level of the individual participants that would explain the interpersonal differences identified with regard to responses to deviations. Possibly the development and use of a more task-specific measure of occupational self-efficacy (cf. Eden & Kinnar, 1991) could have highlighted more differences between the participants. Factors such as individual preferences with regard to task planning, and opinions on how appealing different types of tasks are, could also play a role in task prioritization and task distribution. As was suggested before, the vignette approach can also be used to collect more information about the recovery steps that would follow the described scenarios, than merely when the first recovery step after detection would start, and who would be involved in the newly added recovery tasks. In such studies, the researchers could ask a participant or team of participants (cf. Crew Resource Management (CRM) training in aviation, see for example Helmreich et al., 1999) to take them through all the steps that would be taken, enabling a comparison of the complete responses of different (groups of) participants to the same scenario. During the participants' descriptions, when needed, the researchers can ask for further information or clarification. What is more, it would create the possibility for the researchers to add additional influences at later recovery process stages, for example by taking away available recovery strategies, changing available resources, adding additional tasks or complications.

Future research could also focus on the factors, in addition to those related to task prioritization and distribution and others that have been studied already, that influence recovery processes, separate steps or process phases, transitions between phases, time spent on separate steps and the entire process, and especially the process outcomes: Why do recovery processes turn out the way they do? Insight into those recovery influencing factors would allow assessment of an organization's current recovery possibilities and provide specific leads for possible improvements. Kanse & Van der Schaaf (2001a) and Bove (2002) have provided some initial insights into factors that influence recovery, in both cases based on a literature review supplemented with empirical data. These insights have links with what, for example, Rasmussen (1982) and Swain and Guttmann (1983) (see also Reason, 1990; and Kirwan, 1994) have noted with regard to Performing Shaping Factors (PSFs) that influence the likelihood of errors occurring. Recovery processes are not only influenced by factors characterizing the situation or context in which the process takes place, like errors are, but also by characteristics of the deviation that needs to be recovered. As an example, to make predictions about a recovery process, it would be useful to know how familiar the deviation that needs to be recovered is, how easily observable, traceable (the causes) and reversible it is, how well prepared one is for that situation, and how much effort it will take to recover. Even among deviations that are all the result of failures of the same type, i.e. either technical, or organizational, or human, there might be different answers possible to these questions. Bove (2002) suggested that the factors that influence recovery might be different in different domains. What should be noted too, as was pointed out by Kanse & Van der Schaaf (2001a), is that factors that influence the likelihood of an error do not necessarily also affect recovery from it. Very different factors may play a role in the two processes. In addition, some factors may have an indirect effect on recovery through some other factor. For example, structures provided by the organization may influence team processes or communication between people, and the latter two factors in turn influence recovery.

Clearly, a lot remains to be investigated with regard to failure recovery processes. I hope that with the research presented in this thesis I have not only succeeded in adding to the existing insights into recovery process mechanisms and behavior, but also managed to convince both practitioners and researchers in safety management of the importance of recovery for an organization’s safety and reliability, and inspired some of them to start additional research in this domain.

References

Allwood, C.M. (1984). Error detection processes in statistical problem solving. *Cognitive Science, 8*, 413-437.

Allwood, C.M., & Montgomery, H. (1982). Detection of errors in statistical problem solving. *Scandinavian Journal of Psychology, 23*(2), 131-139.

Argyris, C. (1999). *On organizational learning (2nd edition).* Oxford: Blackwell Publishers.

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technologies and human error* (pp. 271-283). Chichester: John Wiley & Sons.

Bagnara, S., Ferrante, D., Rizzo, A., & Stablum, F. (1988). Causal analysis in error detection and recovery: when does it occur? In *Proceedings of the international conference on joint design of technology, organization and people growth.* October 12-14, 1988, Venice, Italy.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman and Company.

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Beach, L.R., & Mitchell, T.R. (1987). Image theory: Principles, goals and plans. *Acta Psychologica, 66*(3), 201-220.

Bove, T. (2002). *Development and validation of a human error management taxonomy in air traffic control.* PhD Thesis, Risø National Laboratory & University of Roskilde, Roskilde, Denmark.

Bove, T., & Andersen, H.B. (2001). Types of Error Recovery in Air Traffic Management. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, Volume Five, Industrial Ergonomics, HCI, and Applied Cognitive Psychology* (pp. 171-180). Aldershot: Ashgate.

Breakwell, G.M., & Wood, P. (1995). Diary techniques. In G.M. Breakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology*. London: Sage Publications.

Brief, A.P., & Downey, H.K. (1983). Cognitive and organizational structures: A conceptual analysis of implicit organizing theories. *Human relations, 36*(12), 1065-1090.

Brinkman, J.A. (1990). *The analysis of fault diagnosis tasks: Do verbal reports speak for themselves?* PhD thesis, Eindhoven University of Technology.

Brodbeck, F.C., Zapf, D., Prümper, J., & Frese, M. (1993). Error handling in office work with computers: A field study. *Journal of Occupational and Organizational Psychology, 66*(4), 303-317.

Carr, C. (1992). PSS Help when you need it. *Training and Development, 46*(6), 30-38.

Carroll, J.S., & Johnson, E.J. (1990). *Decision research: A field guide.* Newbury park, CA: Sage.

Chen, G., Gully, S.M, & Eden, D. (2001). Validation of a new generalized self-efficacy scale. *Organizational Research Methods, 4*(1), 62-83.

Claessens, B.J.C., Van Eerde, W., Rutte, C.G. & Roe, R.A. (2003). *Planning and completing daily work tasks.* Manuscript submitted for publication.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37-46.

Cohen, M.S., & Freeman, J.T. (1997). Improving critical thinking. In R. Flin, E. Salas, M. Strub, & L. Martin, *Decision making under stress; emerging themes and applications* (pp. 161-169). Aldershot: Ashgate.

Covey, S.R., Merrill, A.R., & Merrill, R.R. (1994). *First things first.* New York: Simon & Schuster.

Cox, S.J., & Tait, N.R.S. (1991). *Reliability, safety & risk management: An integrated approach.* Oxford: Butterworth-Heinemann.

Curry, R.E. (1981). A model of human fault detection for complex dynamic processes. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 171-184). New York: Plenum Press.

Curry, R.E., & Ephrath, A.R. (1976). Monitoring and control of unreliable systems. In T.B. Sheridan & G. Johannsen (Eds.), *Monitoring behavior and supervisory control* (pp. 193-204). New York: Plenum Press.

Curry, R.E., & Gai, E.G. (1976). Detection of random process failures by human monitors. In T.B. Sheridan & G. Johannsen (Eds.), *Monitoring behavior and supervisory control* (pp. 205-220). New York: Plenum Press.

Doireau, P., Wioland, L., & Amalberti, R. (1997). La détection des erreurs humaines par des opérateurs extérieurs à l'action: Le cas du pilotage d'avion. *Le Travail Humain, 60*(2), 131-153.

Dormann, T., & Frese, M. (1994). Error Training: Replication and the function of exploratory behavior. *International Journal of Human-Computer Interaction, 6*(4), 365-372.

Dörner, D, & Wearing, A.J. (1995). Complex problem solving: Toward a (computersimulated) theory. In P.A. Frensch & J. Funke (Eds.), *Complex problem solving: The European Perspective* (pp. 65-102). Hillsdale, NJ: Lawrence Erlbaum Associates.

Duncan, K.D. (1981). Training for fault diagnosis in industrial process plant. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 553-573). New York: Plenum Press.

Duncan, K.D. (1987a). Fault diagnosis training for advanced continuous process installations. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error* (pp. 209-221). Chichester: John Wiley & Sons.

Duncan, K.D. (1987b). Reflections on fault diagnostic experience. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error* (pp. 261-270). Chichester: John Wiley & Sons.

Eden, D., & Kinnar, J. (1991). Modeling Galatea: Boosting self-efficacy to increase volunteering. *Journal of Applied Psychology, 76*(6), 770-780.

Edmondson, A.C. (1996). Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *Journal of Applied Behavioural Science, 32*(1), 5-28.

Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532-550.

Elzer, P.F., Weisang, C., & Zinser, K. (1993). Alarm filtering vs. failure prediction: How to best reduce operator overload. In H.G. Stassen (Ed.), *Analysis, design and evaluation of Man-Machine Systems 1992* (pp. 183-187). Oxford: Pergamon Press.

Embrey, D.E. & Lucas, D.A. (1988). The nature of recovery from error. In L.H.J. Goossens (Ed.), *Human recovery: Proceedings of the COST A1 Seminar on Risk Analysis and Human Error.* Delft: Delft University of Technology.

Farmer, E., Van Rooij, J., Riemersma, J., Jorna, P., & Moraal, J. (Eds.) (1999). *Handbook of Simulator-Based Training.* Aldershot: Ashgate.

Flanagan, J.C. (1954). The Critical Incident Technique. *Psychological Bulletin, 51*, 327-358.

Flin, R., Salas, E., Strub, M., & Martin, L. (Eds.) (1997). *Decision making under stress: Emerging themes and applications.* Aldershot: Ashgate.

Frensch, P.A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P.A. Frensch & J. Funke (Eds.), *Complex problem solving: The European Perspective* (pp. 3-26). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frese, M. (1991). Error management or error prevention: Two strategies to deal with errors in software design. In H.J. Bullinger (Ed.), *Human aspects in computing: Design and use of interactive systems and work with terminals* (pp. 776-782). Amsterdam: Elsevier Science Publishers.

Frese, M.; Brodbeck, F.C.; Zapf, D. & Prümper, J. (1990). The effects of task structure and social support on user's error and error handling. In D. Diaper et al. (Eds.): *Human-Computer Interaction – INTERACT '90* (pp. 35-41). Amsterdam: Elsevier Science Publishers.

Geen, R.G. (1995). *Human motivation: a social psychological approach.* Pacific Grove, CA: Brooks/Cole Publishing Company.

Getzels, J.W. (1982). The problem of the problem. In R.M. Hogarth (Ed.), *Question framing and response consistency*. San Francisco: Jossey Bass.

Grau, J.Y., Doireau, P., & Poisson, R. (1998). Design and use of simulation for training: Lessons drawn from present military use. *Le Travail Humain, 61*(4), 361-385.

Green, S.B., Salkind, N.J., & Akey, T.M. (2000). *Using SPSS for Windows: Analyzing and understanding data.* Upper Saddle River, NJ: Prentice Hall.

Griffin, M.A., Mathieu, J.E., & Jacobs, R.R. (2001). Perceptions of work contexts: Disentangling influences at multiple levels of analysis. *Journal of Occupational and Organizational Psychology, 74*(5), 563-579.

Hale, A.R., & Glendon, A.I. (1987). *Individual behaviour in the control of danger.* Industrial Safety Series Volume 2. Amsterdam: Elsevier.

Hale, A., Wilpert, B., & Freitag, M. (1997). *After the event: From accident to organisational learning.* Oxford: Pergamon.

Heckhausen, H. & Kuhl, J. (1985). From wishes to action: the dead ends and short cuts on the long way to action. In M. Frese & J. Sabini (Eds.), *Goal directed behavior: the concept of action in psychology* (pp. 134-159). Hillsdale, NJ: Lawrence Erlbaum Associates.

Helmreich, R.L., Merrit, A.C., & Wilhelm, J.A. (1999). The evolution of Crew Resource Management training in commercial aviation. *International Journal of Aviation Psychology, 9*(1), 19-32.

Henley, E.J., & Kumamoto, H. (1981). *Reliability engineering and risk assessment.* Englewood Cliffs, NJ: Prentice-Hall.

Hollnagel, E. (1999). Accidents and barriers. In *Proceedings from CSAPC '99, Seventh European Conference on Cognitive Science Approaches to Process Control* (pp.175-180). September 21-24, 1999, Villeneuve d'Asq, France.

Hollnagel, E. (2002). Time and time again. *Theoretical Issues in Ergonomics Science, 3*(2), 143-158.

Hox, J.J., Kreft, I.G.G., & Hermkens, P.L.J. (1991). The analysis of factorial surveys. *Sociological Methods & Research, 19*(4), 493-510.

Hoyos, C.G., & Zimolong, B. (1988). *Occupational safety and accident prevention: Behavioral strategies and methods.* Amsterdam: Elsevier.

Huber, G.P. (1996). Organizational learning: The contributing processes and the literatures. In M.D. Cohen and L.S. Sproull (Eds.), *Organizational learning* (pp. 124-162). Thousand Oaks, CA: Sage.

Hutchins, E. (1996). *Cognition in the wild.* Cambridge, MA: The MIT Press.

Institute of Medicine (in press: November 2003). Patient safety: A new standard for care. Washington, D.C.: The National Academies Press.

Johannsen, G. (1988). Categories of human operator behaviour in fault management situations. In L. Goodstein, M. Andersen & S. Olsen (Eds.), *Task errors and mental models* (pp. 251-258). London: Taylor & Francis.

Kanse, L. (2002). How people recover from failures in risky operations. In R. Kabi (Ed.), *Proceedings 5th International Workshop on Human Error, Safety and Systems Development*, June 17-18, 2002, Newcastle, NSW, Australia.

Kanse, L., & Van der Schaaf, T.W. (2000a). Failure Recovery in Process Industry - An Empirical Comparison of Two Models. In P.C. Cacciabue (Ed.), *Proceedings EAM 2000, 19th European Annual Conference on Human Decision Making and Manual Control* (pp. 153-163), June 26-28, 2000, Ispra, Italy.

Kanse, L., & Van der Schaaf, T.W. (2000b). Recovery from failures – Understanding the positive role of human operators during incidents. In D. de Waard, C. Weickert, J. Hoonhout & J. Ramaekers (Eds.), *Human-System Interaction: Education, research and application in the 21st century. Proceedings Human Factors and Ergonomics Society Europe Chapter Annual Meeting 2000* (pp. 367-379), November 1-3, 2000, Maastricht, the Netherlands. Maastricht: Shaker Publishing.

Kanse, L., & Van der Schaaf, T.W. (2000c). Toepassing van ICT bij het herstellen van fouten (Application of ICT to support recovery from failures). *Gedrag en Organisatie, 13*(6)*,* 360-374.

Kanse, L., & Van der Schaaf, T.W. (2001a). Factors influencing recovery from failures. In R. Onken (Ed.), *Proceedings CSAPC '01, 8th Conference on Cognitive Science Approaches to Process Control,* September 24-26, 2001, Universitat der Bundeswehr, Neubiberg, Germany.

Kanse, L., & Van der Schaaf, T.W. (2001b). Human recovery of failures in the chemical process industry. In D. Harris (Ed.), *Proceedings Third International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 323-332), October 25-27, 2000, Edinburgh, UK. Aldershot: Ashgate.

Kanse, L., & Van der Schaaf, T.W. (2001c). Recovery from failures in the chemical process industry. *International Journal of Cognitive Ergonomics, 5*(3), 199-211.

Kanse, L., Van der Schaaf, T.W., & Rutte, C.G. (2003). A failure has occurred. Now what? Manuscript submitted for publication.

Kaplan, R.S., & Norton, D.P. (1996). *Translating strategy into action: The balanced scorecard.* Boston, MA: Harvard Business School Press.

Kecklund, L.J., & Svenson, O. (1997). Human errors and work performance in a nuclear power plant control room: associations with work-related factors and behavioral coping. *Reliability Engineering and System Safety, 56*(1), 5-15.

Kepner, C.H., & Tregoe, B.B. (1981). *The new rational manager.* Princeton, NJ: Kepner-Tregoe Inc.

Kirwan, B. (1994). *A Guide to Practical Human Reliability Assessment*. London: Taylor & Francis.

Klein, G.A, Orasanu, J., Calderwood, R., Zsambok, C. (Eds.) (1993). *Decision making in action: models and methods.* Norwood, NJ: Ablex Publishing Corporation.

Klein, G.A. (1993). A Recognition-Primed Decision (RPD) Model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: models and methods* (pp. 138-147). Norwood, NJ: Ablex Publishing Corporation.

Klein, G.A. (1997). The Recognition-Primed Decision Model: Looking back, looking forward. In C.E. Zsambok, & G.A. Klein (Eds.), *Naturalistic Decision Making* (pp.285-292). Mahwah, NJ: Lawrence Erlbaum Associates.

Koch, C.J., & Kleinmann, M. (2002). A stitch in time saves nine: Behavioral decision-making explanations for time management problems. *European Journal of Work and Organizational Psychology, 11*(2), 199-127.

Kohn, L.T., Corrigan, J.M., & Donaldson, M.S. (Eds.) (2000). *To err is human: Building a safer health system.* Washington, DC: National Academy Press.

Kontogiannis, T. (1997). A framework for the analysis of cognitive reliability in complex systems: a recovery centred approach. *Reliability Engineering and System Safety, 58*(3), 233-248.

Kontogiannis, T. (1999). User strategies in recovering from errors in man-machine systems. *Safety Science, 32*(1), 49-68.

Koornneef, F. (2000). *Organised learning from small-scale incidents.* PhD thesis, Delft University of Technology.

Kostopoulou, O., & Duncan, K.D. (2001). Abstract and reduced-context representations in fault-finding training. *Ergonomics, 44*(2), 175-201.

Kroemer, K.H.E., Kroemer, H.B., & Kroemer-Elbert, K.E. (2001). *Ergonomics: How to design for ease and efficiency (2nd edition.).* Upper Saddle River, NJ: Prentice-Hall.

Kumamoto, H., & Henley, E.J. (1996). *Probabilistic risk assessment and management for engineers and scientists (2nd edition).* New York: IEEE Press.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lawton, R. (1998). Not working to rule: Understanding procedural violations at work. *Safety Science, 28*(2), 77-95.

Liker, J.K. (1982). Family prestige judgments. In P.H. Rossi & S.L. Nock (Eds.), *Measuring social judgments* (pp.119-144). Beverly Hills: Sage Publications.

Lipshitz, R.,& Bar-Ilan, O. (1996). How problems are solved: Reconsidering the Phase Theorem. *Organizational Behavior and Human Decision Processes, 65*(1), 48-60.

Marshall, E.C., & Shepherd, A. (1981). A fault-finding programme for continuous plant operators. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 575-588). New York: Plenum Press.

Mason, M.A., & Redmon, W.K. (1992). Effects of immediate versus delayed feedback on error detection accuracy in a quality control simulation. *Journal of Organizational Behavior Management, 13*(1), 49-83.

Maule, A.J. (1997). Strategies for adapting to time pressure. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: Emerging themes and applications* (pp.71-279). Aldershot: Ashgate.

McGraw, K.L. (1994). Performance support systems: Integrating AI, hypermedia, and CBT to enhance user performance. *Journal-of-Artificial-Intelligence-in-Education, 5*(1), 3-26.

Meacham, J.A., & Cooney Emont, N. (1989). The interpersonal basis of everyday problem solving. In J.D. Sinnott (Ed.), *Everyday problem solving: Theory and applications* (pp. 7-23). New York: Praeger.

Mehl, K., & Schuette, M. (1999). Simulators – A perspective on what to train and what to analyze regarding human reliability. In Schuëller & Kafka (Eds.), *Safety and Reliability* (pp. 675-680). Rotterdam: Balkema.

Millot, P., Taborin, V., Kamoun, A., & Willaeys, D. (1988). Effects of the dynamic allocation of supervision tasks between man and computer on the performance of automated processes. In J. Patrick & K.D. Duncan (Eds.), *Training, human decision making and control* (pp. 175-187). Amsterdam: Elsevier Science Publishers.

Mintzberg, H. (1983). *Structures in fives.* Englewood Cliffs, NJ: Prentice-Hall.

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of 'unstructured' decision processes. *Administrative Science Quarterly, 21*, 246-275.

Mo, J., & Crouzet, Y. (1996). Human error tolerant design for air-traffic control systems. In *Proceedings of the third Probability Safety Assessment and Management Conference, PSAM-III*, June, 1996, Crete, Greece.

Montgomery, H. (1983). Decision rules and the search for a dominance structure: Towards a process model of decision making. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Advances in psychology*. Amsterdam: North-Holland.

Moray, N. (1981). The role of attention in the detection of errors and the diagnosis of failures in man-machine systems. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 185-198). New York: Plenum Press.

Morrison, D.L., & Duncan, K.D. (1988). Strategies and tactics in fault diagnosis. *Ergonomics, 31*(5), 761-784.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Noble, D. (1989). *Application of a theory of cognition to situation assessment.* Vienna, VA: Engineering Research Associates.

Norman, D.A. (1981). Categorization of action slips. *Psychological Review, 88*(1), 1-23.

Norman, D.A. (1984). Stages and levels in human-machine interaction. *International Journal of Man-Machine Studies, 21*(4), 365-375.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory.* New York: McGraw-Hill.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: models and methods* (pp. 3-20). Norwood, NJ: Ablex Publishing Corporation.

Parasuraman, R., Molloy, R., & Singh, I.L. (1993). Performance consequences of automation-induced complacency. *International Journal of Aviation Psychology, 3*(1), 1-23.

Patrick, J., & Stammers, R.B. (1981). The role of computers in training for problem diagnosis. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 589-604). New York: Plenum Press.

Perrow, C. (1984). *Normal accidents: living with high-risk technology*. New York: Basic Books.

Pliske, R, & Klein, G. (2003). The naturalistic decision-making perspective. In S.L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 559-585). Cambridge, UK: Cambridge University Press.

Pritchard, R.D. (1990). *Measuring and improving organizational productivity: A practical guide*. New York: Praeger.

Rabbitt, P.M. (1978). Detection of errors by skilled typists. *Ergonomics, 21*(11), 945-958.

Rasbash, J., Browne, W., Healy, M., Cameron, B., & Charlton, C. (2002). *MLwiN test version 1.2.0001. Multilevel models project.* London: Institute of Education.

Rasmussen, J. (1976). Outlines of a hybrid model of the process operator. In T.B. Sheridan and G. Johannsen (Eds.), *Monitoring behavior and supervisory control* (pp. 371-383). New York: Plenum Press.

Rasmussen, J. (1982). Human Errors. A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4,* 311-333.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* New York: Elsevier Science Publishing Co.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science, 27*(2/3), 183-213.

Reason, J. (1987). Generic Error-Modelling System (GEMS): A Cognitive Framework for Locating Common Human Error Forms. In J. Rasmussen, K. Duncan & J. Leplat (Eds.), *New Technology and Human Error* (pp. 63-83). Chichester: John Wiley & Sons.

Reason, J. (1990). *Human Error.* Cambridge, UK: Cambridge University Press.

Reason, J. (1991). Too little and too late: a commentary on accident and incident reporting systems. In T.W. van der Schaaf, D.A. Lucas & A.R. Hale (Eds.), *Near miss reporting as a safety tool* (pp. 9-26). Oxford: Butterworth Heinemann.

Reason, J. (1997). *Managing the risk of organisational accidents.* Hampshire: Ashgate.

Reason, J. & Lucas, D. (1984). Using cognitive diaries to investigate naturally occurring memory blocks. In J.E. Harris & P.E. Morris (Eds.), *Everyday memory, actions and absent-mindedness* (pp. 53-70). London: Academic Press.

Reason, J., & Mycielska, K. (1982). *Absent-Minded? The psychology of mental lapses and everyday errors.* Englewood Cliffs, NJ: Prentice-Hall.

Rizzo, A., Bagnara, S., & Visciola, M. (1987). Human error detection processes. *International Journal of Man-Machine Studies, 27*(5&6), 555-570.

Rizzo, A., Ferrante, D., & Bagnara, S. (1995). Handling human errors. In J.M. Hoc, P.C. Cacciabue & E. Hollnagel (Eds.), *Expertise and Technology: Cognition and Human Computer Interaction* (pp. 195-212). Hillsdale, NJ: Lawrence Erlbaum Associates.

Roberts, K.H. & Bea, R.G. (2001). Must accidents happen? Lessons from high-reliability organizations. *Academy of Management Executive, 15*(3), 70-77.

Rossi, P.H., & Anderson, A.B. (1982). The factorial survey approach: An introduction. In P.H. Rossi & S.L. Nock (Eds.), *Measuring social judgments* (pp. 15-69). Beverly Hills: Sage Publications.

Rossi, P.H., & Nock, S.L. (Eds.) (1982). *Measuring social judgments.* Beverly Hills: Sage Publications.

Rouse, W.B. (1981). Experimental studies and mathematical models of human problem solving performance in fault diagnosis tasks. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 199-216). New York: Plenum Press.

Rouse, W.B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30*(4), 431-443.

Salas, E., & Klein, G. (Eds.) (2001). *Linking expertise and naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.

Salvendy, G. (Ed.). (1997). *Handbook of human factors and ergonomics* (2nd edition). Chichester: Wiley-Interscience.

Sarter, N.B., & Alexander, H.M. (2000). Error types and related error detection mechanisms in the aviation domain: An analysis of aviation safety reporting system incident reports. *The international journal of aviation psychology, 10*(2), 189-206.

Schyns, B., & Von Collani, G. (2002). A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *European Journal of Work and Organizational Psychology, 11*(2), 219-241.

Sellen, A.J. (1994). Detection of Everyday Errors. *Applied Psychology: An International Review, 43*(4), 475-498.

Shelton, S.H. (1990). Developing the construct of general self-efficacy. *Psychological Reports, 66*(3), 987-994.

Shepherd, A. (1986). Issues in the training of process operators. *International Journal of Industrial Ergonomics, 1*(1), 49-64.

Sherer, M., Maddux, J.E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports, 51*(2), 663-671.

Sheridan, T.B. (1997). Supervisory control. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1295-1327). Chichester: Wiley-Interscience.

Simon, H.A. (1960). *The new science of management decision.* New York: Harper.

Snijders, T.A.B., & Bosker, R.J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage Publications.

Stanton, N.A., & Baber, C. (1995). Alarm-initiated activities: An analysis of alarm handling by operators using text-based systems in supervisory control systems. *Ergonomics, 38*(11), 2414-2431.

Svenson, O. (1991). The Accident Evolution and Barrier function (AEB) model applied to incident analysis in the processing industries. *Risk Analysis, 11*(3), 499-507.

Svenson, O. (2001). Accident and incident analysis based on the accident evolution and barrier function (AEB) model. *Cognition, Technology & Work, 3*(1), 42-52.

Swain, A.D. & Guttmann, H.E. (1983). *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications.* NUREG/CR 1278. Albuquerque, NM: Sandia National Laboratories.

Symon, G. (1998). Qualitative research diaries. In G. Symon & C. Cassell (Eds.), *Qualitative methods and analysis in organizational research: A practical guide* (pp. 94-117). London: Sage Publications.

Turner, B.A., & Pidgeon, N.F (1997). *Man-Made Disasters (2nd edition)*. Oxford: Butterworth-Heinemann.

Van der Schaaf, T.W. (1988). Critical incidents and human recovery: Some examples of research techniques. In L.H.J. Goossens (Ed.), *Human recovery: Proceedings of the COST A1 Seminar on Risk Analysis and Human Error.* Delft: Delft University of Technology.

Van der Schaaf, T.W. (1992). *Near miss reporting in the chemical process industry.* PhD thesis, Eindhoven University of Technology.

Van der Schaaf, T.W. (1995). Human recovery of errors in man-machine systems. In *Proceedings of the 6th IFAC Symposium on Analysis, Design, and Evaluation of Man-Machine Systems*, June 27-29, 1995, MIT, Cambridge, MA, USA.

Van der Schaaf, T.W. (1996). PRISMA: A Risk Management Tool Based on Incident Analysis. In *Proceedings of the International Conference and Workshop on Process Safety Management and Inherently Safer Processes* (pp. 242-251). October 8-11, 1996, Orlando, Florida.

Van der Schaaf, T.W., & Kanse, L. (1999). Error recovery in socio-technical systems. In J.M. Hoc, P. Millot, E. Hollnagel & P.C. Cacciabue (Eds.), *Proceedings CSAPC '99, Seventh European Conference on Cognitive Science Approaches to Process Control* (pp.151-156), September 21-24, 1999, Villeneuve d'Asq, France.

Van der Schaaf, T.W., & Kanse, L. (2000). Errors and error recovery. In P.F. Elzer, R.H. Kluwe & B. Boussoffara (Eds.), *Human Error and System Design and Management* (pp. 27-38). London: Springer Verlag.

Van der Schaaf, T.W., & Kanse, L. (2004). Biases in incident reporting databases: An empirical study in the chemical process industry. *Safety Science, 42*(1), 57-67.

Van der Schaaf, T.W., Lucas, D.A., & Hale, A.R. (Eds.) (1991). *Near miss reporting as a safety tool*. Oxford: Butterworth Heinemann.

Van Dyck, C. (2000). *Putting errors to good use: Error management culture in organizations.* PhD thesis, Amsterdam: KLI.

Van Vuuren, W. & Van der Schaaf, T.W. (1995). Modelling organisational factors of human reliability in complex man-machine systems. In *Proceedings of the 6th IFAC Symposium on Analysis, Design and Evaluation of Man-Machine Systems* (pp. 323-328). June 27-29, 1995, M.I.T., Cambridge, MA, USA.

Van Vuuren, W. (1998). *Organisational failure: an exploratory study in the steel industry and the medical domain*. PhD thesis, Eindhoven University of Technology.

Van Vuuren, W. (1999). Organisational failure: lessons from industry applied in the medical domain. *Safety Science, 33* (1-2), 13-29.

Van Vuuren, W., & Kanse, L. (in press). Widening the scope of incident analysis in complex work environments. In T. Manser (Ed.), *Komplexes Handeln in der Anaesthesie (Complex Activity in Anesthesiology)* (pp. 186-207). Berlin: Pabst Science Publishers.

Wagenaar, W.A., & Groeneweg, J. (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies, 27*(5-6), 587-598.

Waller, M.J. (1997). Keeping pins in the air: How work groups juggle multiple tasks. *Advances in Interdisciplinary Studies of Work Teams, 4*, 217-247.

Waller, M.J. (1999). The timing of adaptive group responses to nonroutine events. *Academy of Management Journal, 42*(2), 127-137.

Weick, K.E., Sutcliffe, K.M., & Obstfeld, D. (1999). Organizing for high reliability: Processes of collective mindfulness. In B.M. Staw & R. Sutton (Eds.), *Research in organizational behavior, volume 21* (pp. 81-123). Greenwich, CT: JAI Press.

Wickens, C.D., Gordon, S.E., & Liu, Y. (1998). *An introduction to human factors engineering.* New York: Longman.

Wickens, C.D., & Kessel, C. (1981). Failure detection in dynamic systems. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 155-170). New York: Plenum Press.

Wioland, L., & Amalberti, R. (1998). *Human error management: towards an ecological safety model a case study in an air traffic control microworld.* Paper presented at the 9th European conference on cognitive ergonomics. Limerick University, Ireland, August 24-26, 1998.

Yin, R.K. (1994). *Case study research: Design and methods* (2^{nd} edition). Thousand Oaks, California: Sage Publications.

Zapf, D., Maier, G.W., Rappenberger, G., & Irmer, C. (1994). Error detection, task characteristics, and some consequences for software design. *Applied Psychology: An International Review, 43*(4), 433-453.

Zapf, D. & Reason, J.T. (1994). Introduction: Human errors and error handling. *Applied Psychology: An International Review, 43*(4), 427-432.

Zsambok, C.E., & Klein, G. (Eds.) (1997). *Naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix 1 Glossary

accident	event where a *failure* or combination of failures has eventually led to at least some undesired safety-, reliability-, or otherwise performance-related *negative consequences*
causal tree analysis	*incident* analysis process whereby one systematically traces how an incident (the tree's top event) was caused by a number of factors which each on their turn have their own causes, and so forth, until the so-called *root causes* of the incident are identified (usually more than one), for which no further explanation can be found within the organization or context in which the incident occurred; this process allows for the inclusion of information about the *recovery steps* taken (if any)
countermeasures (C)	*recovery process phase* involving the planning and implementation of actions to return the situation to normal or at least to limit the consequences
detection (D)	*recovery process phase* in which a deviation is detected
deviation	problem situation immediately resulting from a *failure* or combination of failures, which is considered to be an unacceptable variation from the normal, intended situation; this is how these failures manifest themselves before actually leading to their associated potential *negative consequences*
error	see *human failure*
explanation (E)	*recovery process phase* in which people look for further information about the *deviation* and its causes
factorial object	see *vignette*

failure	a falling short of what was expected; depending on its origin, different types can be distinguished: *technical*, *organizational*, or *human*
human failure	synonymous with error; *failure* committed by the people at operational levels in an organization; among them, based on the behavior- or performance level at which they occur, *skill-, rule-, and knowledge-based* errors can be distinguished
incident	event where a *failure* or combination of failures has occurred with the potential to lead to negative safety-, reliability-, or otherwise performance-related consequences, irrespective of whether in the end these *negative consequences* became manifest, at least to some extent, or were avoided completely; this definition encompasses both *accidents* and *near misses*
knowledge-based behavior	occurs when actions are performed that are novel, in unfamiliar situations, involves a substantial amount of reasoning
NDM	Naturalistic Decision Making
near miss	event where undesired *negative consequences* have been avoided, either because all the necessary *recovery steps* have been performed in a timely and effective manner, or because of sheer luck or coincidence
negative consequences	the potential, undesired end result of a *failure* or combination of failures, that do not take place in the case of a *near miss*, but that occur, at least to some extent, in the case of an *accident*
occupational self-efficacy	*self-efficacy* within occupational, i.e. work-related context

organizational failure	*failure* occurring at the management levels in an organization; these failures most often lead indirectly to incidents by triggering other failures
recovery from self-made errors	recovery from *errors* made by the same people as those involved in the subsequent *recovery process*, as opposed to recovery from somebody else's errors
recovery process	the detection that something is wrong (detection of the *deviation*) and all the information processing and actions performed in response to this to avoid *negative consequences*, until the situation is returned to normal; this process involves a sequence of *steps* or *phases*
recovery process phase	a phase in the sequence of *recovery steps*, for which a specific goal can be distinguished, i.e. either *detection*, *explanation*, or *countermeasures*; this term is a synonym for recovery step
recovery step	any action or cognitive process or combination thereof that forms part of a *recovery process*, can be categorized based on its goal as belonging to a specific *recovery process phase*; this term is used interchangeably with recovery process phase
root cause	factor at the basis of an incident's *causal tree*, which in combination with the other root causes has led to the occurrence of the *incident*; for root causes, no further underlying causes can be identified within the organization of context in which the incident occurred
rule-based behavior	occurs when actions are performed which can be selected from a repertoire of predefined actions, governed by rules that are still relatively familiar
self-efficacy	a person's conviction that he or she can successfully execute a given behavior required to produce certain outcomes

skill-based behavior	occurs when actions are performed almost without thinking, automatically
SRK framework/model	framework in which different information processing behavior or performance levels are distinguished based on the amount of conscious reasoning involved; ordered from very little to a lot of conscious reasoning involved, these levels are *skill-based, rule-based, and knowledge-based behavior*
technical failure	*failure* occurring in technical systems of an organization; related to the design or the construction of technical systems, installations, machinery, tools and equipment, or simple material or component defects
vignette	description of a situation, focused on a few situation characteristics, that is presented to a respondent who is asked a number of questions regarding this situation, synonymous with *factorial object*

Appendix 2 Example event diary proforma

Event reporting form (originally 2 pages A4 size)
for study of recovery from self-made errors

Date on which event took place?	
Morning, afternoon, or night shift?	

	Short description of the error	Time at which error was made	Circumstances under which this error was made
Describe the error you made			

What could have happened if no recovery steps would have been taken?	
How soon could that happen?	

	Short description of how it was discovered that something was wrong	Who detected that something was wrong	Time at which this deviation was detected	Circumstances that helped / hindered the detection of the deviation
If someone else warned you that something was wrong, how did this person detect this? *Answer only if applicable!*				
How did you (consequently) detect that something was wrong?				

see over

	Describe recovery step taken	Goal of the recovery step? (circle the applicable answer: A = analysis of deviation and its causes; IC = immediate correction; LC = long term correction)	By whom was this step taken	Start time	Duration	Circumstances that helped / hindered this step
Describe the recovery process that followed the detection that something was wrong step by step in the spaces provided here:		A / IC / LC				
		A / IC / LC				
		A / IC / LC				
		A / IC / LC				
		A / IC / LC				

Describe consequences of the error (if any) that remained after all the recovery steps were taken	

Would you normally report this event via the existing near miss reporting system? (circle the answer of your choice)	yes / no / maybe
Why?	

Thank you very much for this event report!

Appendix 3 Vignette examples

Example 1

Recovery from problem situations at work

Imagine yourself in the following situation at work:

You detect a problem in the plant.
- Procedures or unwritten work routines covering this problem allow you to perform the required recovery tasks alone, by yourself.
- You have the required knowledge and expertise.
- The tools, equipment and process information needed for the recovery actions are available.
- The problem can lead to severe consequences if nothing is done about it.
- These consequences can only occur after a longer period of time.
- The entire recovery task takes a lot of time.

An example of such a situation would be:
On your rounds through the plant you discover that the stand-by filter of a pump has not been cleaned.
- You and co-workers in the same job are responsible for cleaning the filter.
- You have done this task often before, you know how to do this.
- The gloves, wrenches and other tools and parts you need are readily available.
- If nothing is done, this situation can lead to a pump trip causing the product flow to stop, which can eventually lead to a plant shutdown.
- This will only happen as soon as the current online filter has to be changed when it has become clogged with dirt, as then both filters are clogged.
- It will take circa one hour to dismantle, clean, and reinstall the stand-by filter.

The questions that follow concern what you would do in the situation described above – not only in the specific example but also more generally **in situations with the characteristics described above the example**. Please mark the answers that correspond most closely with what would happen with a cross (X).

Case A:

The situation described above occurs.
At this moment you are **busy** with other urgent and important tasks.
Others are available who can assist or take over tasks.

		entirely by myself	mostly self, little bit by others	half self, half by others	little bit self, mostly by others	entirely by others
1	To what extent would you involve others in the recovery tasks?	☐	☐	☐	☐	☐

		immediately	later during the same shift	in a later shift
2	When does the first recovery action start (irrespective of by whom)?	☐	☐	☐

Case B:

The situation described above occurs.
At this moment you are **busy** with other urgent and important tasks.
No others are available who can assist or take over tasks.

		entirely by myself	mostly self, little bit by others	half self, half by others	little bit self, mostly by others	entirely by others
1	To what extent would you involve others in the recovery tasks?	☐	☐	☐	☐	☐

		immediately	later during the same shift	in a later shift
2	When does the first recovery action start (irrespective of by whom)?	☐	☐	☐

Case C:

The situation described above occurs.
The other tasks on which you are already working are **not so urgent and not so important**.
Others are available who can assist or take over tasks.

1	To what extent would you involve others in the recovery tasks?	entirely by myself ☐	mostly self, little bit by others ☐	half self, half by others ☐	little bit self, mostly by others ☐	entirely by others ☐

2	When does the first recovery action start (irrespective of by whom)?	immediately ☐	later during the same shift ☐	in a later shift ☐

Case D:

The situation described above occurs.
The other tasks on which you are already working are **not so urgent and not so important**.
No others are available who can assist or take over tasks.

1	To what extent would you involve others in the recovery tasks?	entirely by myself ☐	mostly self, little bit by others ☐	half self, half by others ☐	little bit self, mostly by others ☐	entirely by others ☐

2	When does the first recovery action start (irrespective of by whom)?	immediately ☐	later during the same shift ☐	in a later shift ☐

Thank you for your answers!

Example 2
(the pages with questions that follow the vignette are not included again)

Recovery from problem situations at work

Imagine yourself in the following situation at work:

You detect a problem in the plant.

- Procedures or unwritten work routines covering this problem suggest that the required recovery tasks should be handed over completely to others.
- You have the knowledge and expertise required for the recovery task.
- The tools, equipment and process information needed for the recovery actions are available.
- The problem can lead to severe consequences if nothing is done about it.
- These consequences can occur within a short amount of time.
- The entire recovery task does not take much time to complete.

An example of such a situation would be:
You detect a small fire near the storage tanks in the dry grass that has been mown recently.

- Normally, the on-site fire department has to be called in to deal with this.
- You have been trained in fire fighting and know how to extinguish this fire.
- You are allowed to work with the required equipment and a hydrant with a fire hose is available nearby.
- If nothing is done the fire can spread, become hard to extinguish, and lead to explosion hazard.
- That risk is already present and will only become larger.
- Normally it takes circa twenty minutes to bring over the required equipment and to completely extinguish a fire of this size.

The questions that follow concern what you would do in the situation described above – not only in the specific example but also more generally **in situations with the characteristics described above the example**. Please mark the answers that correspond most closely with what would happen with a cross (X).

Appendix 4 Questionnaire used for person-related factors

Finally, a few more general questions, not connected to the scenarios.

Again, individual answers will not be shown to your organization, and will not be used to identify persons – they are exclusively meant as a means to establish relationships in the research data.

For the following statements, can you please indicate the extent to which you agree or disagree with them, with regard to your current work environment? Just mark the box that corresponds with the answer you find most applicable, with a cross (X).

	I totally disagree	I disagree	I don't agree and don't disagree	I agree	I totally agree
Thanks to my resourcefulness, I know how to handle unforeseen situations in my job.	☐	☐	☐	☐	☐
If I am in trouble at my work, I can usually think of something to do.	☐	☐	☐	☐	☐
I can remain calm when facing difficulties in my job because I can rely on my abilities.	☐	☐	☐	☐	☐
When I am confronted with a problem in my job, I can usually find several solutions.	☐	☐	☐	☐	☐
No matter what comes my way in my job, I'm usually able to handle it.	☐	☐	☐	☐	☐
My past experiences in my job have prepared my well for my occupational future.	☐	☐	☐	☐	☐
I meet the goals that I set for myself in my job.	☐	☐	☐	☐	☐
I feel prepared to meet most of the demands in my job.	☐	☐	☐	☐	☐

How many years of work experience do you have in process operations? .... years

Thank you very much for your cooperation!

Is there anything you would like to let us know with regard to this study? You can write down any comments and suggestions you may have below and on the other side of this page; if applicable, you can also write your comments next to the scenario descriptions.

....
....
....
....
....
....

Summary

Until recently, the focus of safety and reliability management has mainly been on the prevention of failures. However, additional benefits can be gained by promoting the possibilities to recover once failures have occurred, but before they have led to negative consequences. After all, not all failures can be foreseen, and even foreseen failures can not always be prevented – the appropriate preventive measures may be unknown, impossible to implement, or the benefits of implementing them may not outweigh the costs. Although people are often seen as those who cause things to go wrong by committing errors (i.e. human failures), they can also positively contribute to an organization's safety and reliability, through their ability to recover from failures. The objective of the research described in this thesis was to gain a better understanding of how people do this, of the processes they follow, starting with the detection that something is wrong, followed by all the information processing and actions performed in response to this, until the situation is returned to normal. Several aspects of such recovery processes are studied in this thesis.

The research described in this thesis started with a review of relevant literature from the domains of work and organizational psychology, cognitive psychology, human reliability, ergonomics, safety management, problem solving, and (naturalistic) decision making. The review showed that most of the existing recovery-related research focuses on recovery from errors, that is human failures, whereas recovery from other types and combinations of failures has received far less attention. Furthermore, while in the literature a reasonable amount of attention was given to the detection of failures, the available insights regarding the steps that follow detection were not only fewer but also more divergent. Seven more specific research questions resulted from this literature review; the first five focusing on respectively the steps or phases involved, the behavior these require, some time-related aspects, how tasks are prioritized, and the people who play a role; and the last two linking these aspects to the events preceding a recovery process and the process end results.

To answer these research questions, a total of five studies were performed in the chemical process industry. An increasing number of organizations in this industry have set up systems via which near misses are reported, that is, situations where an accident could have occurred but where negative consequences were avoided, most often as a result of successful recovery. The predominant use of such systems has been to identify failures against which preventive measures need to be implemented. However, the reported events can also be used to provide insight into the reasons why they did not develop into accidents, and into the recovery processes that were followed. These additional learning opportunities have remained largely untapped so far.

In the initial three studies that were performed to answer the research questions, a multiple case study approach was used to collect data on a wide variety of recovery processes, in two different chemical plants. In the first study, near miss reports from one chemical plant were used as the basis for data collection, followed by interviews with those involved in the reported events. In the second study, performed in the same plant, event diaries were used to specifically collect data on recovery from self-made errors, since the data collected in the first study did not include any such events. In the third study, performed in another chemical plant, interviews were held in which people were asked to recall and describe recent situations in which they had more or less successfully avoided negative consequences through the implementation of recovery actions. The collected data were coded to allow further analysis of and comparisons between the recovery processes. Data was collected about a total of 122 recovery processes.

The initial studies demonstrated that the recovery processes always started with the detection of a deviation, as per definition, and that after that, any number and combination of steps aimed at explanation and countermeasures were possible, ranging from no such steps occurring at all to multiple recurrences of either one, most often ending with a countermeasures phase. Several subgoals were identified within the explanation phase (definition of the problem and identification of the causes) and the countermeasures phase (stabilization, mitigation, temporary correction, and permanent correction). The studies also demonstrated the important role of ad-hoc and knowledge-based behavior in recovery processes, that is, of people's creative thinking abilities. This indicated the importance for organizations to invest in their employees' detailed process and plant knowledge to prepare them even for the unexpected. Pre-planning was predominantly found in the countermeasures phase and was found to significantly speed up these steps. So, where fast recovery is essential, organizations should provide pre-planned countermeasures. The early steps in recovery processes were shown to be completed the fastest, their goal often being to bring things back under control. The overall time between recovery process start and completion did not only depend on time spent on the recovery steps, but also on the time elapsed in between those steps. Therefore organizations should at least strive to get the potential hazards associated with a deviation contained within as little time as possible, early on in the process, before other tasks are allowed to delay the progress of the recovery process. The presence of other urgent and important tasks was shown to influence prioritization, in that they led to an increase in the total amount of time that elapsed between recovery process start and completion. To ensure that people make the correct decisions in task prioritization, organizations should provide them with the knowledge required to make correct assessments of the urgency and importance of both recovery and normal tasks. The involvement of others was found to be not only important for the detection of deviations, but also for later steps in the recovery process, because of their qualifications, expertise, the equipment they had, or the additional capacity they could offer, or a combination of those reasons. Therefore, organizations should ensure that access to and communication with such others is made as easy as possible. With regard to the relation between the events preceding a recovery process, the process itself, and its end results, a few typical patterns emerged from the research data. First, the presence of technical and organizational failures increased the need for explanation. The presence of organizational failures also delayed the moment of detection. Furthermore, late detection led to longer processes in which more people were involved, such processes also involved more steps, and as result of all of these things combined, more, or more severe consequences remained at the end. Because of this, organizations may want to focus on reducing the time elapsed before detection as much as possible. And finally, cases involving technical and/or organizational failures and/or more, or more severe potential consequences led to longer processes involving more steps and people than cases involving only human failures and/or few or minor potential consequences. As a result of all these things combined, more, or more severe consequences remained after the recovery process. This indicated that there may be some room for improvement regarding the ability to recover from failure combinations involving technical or organizational failures. Nevertheless, care should be taken not to lose present expertise regarding recovery from only human failures, since the potential consequences of such failures if no recovery steps are taken can also be very serious.

Two follow-up studies were performed among operators in two additional chemical plants, to gain a more complete understanding of the factors influencing task prioritization and the distribution of tasks across people in situations where recovery is needed. Contrary to the initial studies, in these follow-up studies, the combination in which these influences on task prioritization and distribution were present was taken into account. Furthermore, whereas in the initial studies for each recovery process data was only available regarding the specific response of those who were actually involved in that process, in the follow-up studies the same scenario of a situation where recovery was needed was presented to several participants. Scenarios, or vignettes, were developed containing both an abstract description of a situation in need of recovery and an example. The values or levels of the factors that were expected to influence task prioritization and distribution varied across these scenarios. A subset of the entire set of vignettes was given to each participant. For each scenario, they had to indicate when the necessary recovery tasks would be started, and to what extent they would involve others in the recovery tasks. Based on the expectation that their answers would not only be influenced by the factors built into the scenarios, but also by some person-related factors, a questionnaire was given to all of the participants, too, to establish the number of years of their relevant work experience and their scores on an occupational self-efficacy scale. The data collected via the vignettes and questionnaires was analyzed using multilevel analysis and model building techniques. By using these techniques, the hierarchical or nested character of the data (vignettes nested in participants) was taken into account. A total of 222 operators participated in the follow-up studies, who together completed 3540 scenarios.

The follow-up studies showed that urgency and importance were important factors in prioritization in situations where recovery from a deviation was needed, which was in line with findings from time management research, action theory, and theory on motivation and goal hierarchies. This finding again underlined the importance for organizations to ensure that their employees make the correct assessments of the urgency and importance of both recovery and normal tasks. The availability of others proved to be a key factor in task distribution, which again showed the importance of access to and communication with such others in situations where their assistance is needed. The nature of the other tasks on which people were already working when they detected the deviation, i.e. how urgent and important those tasks were, also played an important role in task distribution, and so did the availability of the required knowledge and expertise. The follow-up studies demonstrated that in certain circumstances people would deviate from what the normal work routines or procedures prescribed regarding who should be involved in the recovery tasks. To a great extent the availability of others, other tasks, and knowledge and expertise also determined whether such procedure deviations were likely to occur. These procedure deviations were not necessarily a bad thing; the intention of those involved was to get the job done and to avoid negative consequences. Of course, if people deviate from what procedures prescribe or normal work routines suggest, it is vital that those who end up performing tasks that are not theirs know what they are doing, otherwise additional risks arise.

The implications of the findings from the five empirical studies on which this thesis was based are not limited to domain of the chemical process industry. While the exact form recovery processes take, the available time, resources, and so on may be different in other domains, the general process mechanisms and behaviors identified in this thesis, and the described benefits from focusing on recovery still apply. Even though the research described in this thesis has already added to the existing insights into recovery processes, a lot remains to be investigated in this area. For example, it may be particularly interesting to develop specific ways or methods to improve an organization's recovery possibilities and recovery performance, to implement these, and to measure their effectiveness afterwards.

Nederlandse samenvatting

De focus bij het management van proces- en productbetrouwbaarheid, veiligheid en kwaliteit was tot zover voornamelijk gericht op de preventie van fouten. Het kan echter ook erg nuttig zijn om de herstelmogelijkheden zodra er fouten zijn opgetreden, maar voordat deze tot negatieve gevolgen hebben geleid, te optimaliseren. Niet alle fouten kunnen tenslotte worden voorzien, en zelfs voorziene fouten kunnen niet altijd worden voorkomen – de gepaste preventieve maatregelen kunnen onbekend zijn, of het implementeren ervan onmogelijk, of de baten van het implementeren wegen wellicht niet op tegen de kosten. Hoewel mensen vaak gezien worden als degenen die dingen mis laten gaan door fouten te maken (menselijke fouten), kunnen zij ook op een positieve wijze bijdragen aan de veiligheid en betrouwbaarheid in een organisatie, door hun vermogen om fouten te herstellen. Het doel van het in dit proefschrift beschreven onderzoek was om een beter inzicht te krijgen in hoe mensen dit doen, in de herstelprocessen die zij doorlopen, beginnend met de ontdekking dat er iets mis is, gevolgd door alle denkprocessen en handelingen in reactie daarop, totdat de situatie naar normaal is teruggekeerd. Verschillende aspecten van zulke herstelprocessen zijn in dit proefschrift bestudeerd.

Het onderzoek dat in dit proefschrift is beschreven begon met een overzicht van relevante literatuur uit de domeinen van arbeids- en organisatiepsychologie, cognitieve psychologie, menselijk falen en betrouwbaarheid, ergonomie, veiligheidsmanagement, probleemoplossen en besluitvorming. Het overzicht liet zien dat het grootste deel van het bestaande, herstelproces-gerelateerde onderzoek gericht was op het herstellen van menselijke fouten, en er veel minder aandacht was gericht op het herstellen van andere soorten en combinaties van fouten. Bovendien, terwijl in de literatuur redelijk wat aandacht was geschonken aan het ontdekken van fouten, waren bestaande inzichten met betrekking tot de stappen die na zo'n ontdekking volgen niet alleen minder talrijk maar ook meer uiteenlopend. Op basis van dit literatuuroverzicht werden zeven meer specifieke onderzoeksvragen geformuleerd. De eerste vijf vragen waren gericht op respectievelijk de stappen of fasen waaruit het proces bestaat, het soort gedrag dat hiervoor nodig is, enkele tijdgerelateerde aspecten, hoe prioriteiten voor taken vastgesteld worden, en de mensen die een rol spelen. In de laatste twee vragen werden deze aspecten gerelateerd aan de gebeurtenissen voorafgaand aan een herstelproces en de eindresultaten van het proces.

Om de onderzoeksvragen te beantwoorden, werden er in totaal vijf studies uitgevoerd in de chemische procesindustrie. Een groeiend aantal van de organisaties in deze industrie heeft systemen opgezet waarbinnen bijna-ongevallen worden gerapporteerd, dat wil zeggen, situaties waarbij een ongeval had kunnen plaatsvinden maar waar negatieve gevolgen werden voorkomen, meestal als gevolg van succesvol herstel. Zulke meldingssystemen

worden voornamelijk gebruikt om fouten op te sporen waartegen preventieve maatregelen genomen moeten worden. De gerapporteerde gebeurtenissen kunnen echter ook gebruikt worden om inzicht te krijgen in de redenen waarom deze zich niet tot heuse ongevallen ontwikkeld hebben en in de herstelprocessen die gevolgd zijn. Deze additionele leermogelijkheden worden tot op heden nog bijna niet benut.

In de eerste drie studies die uitgevoerd werden om de onderzoeksvragen te beantwoorden, werd een meervoudige case studie aanpak gehanteerd om data te verzamelen over een grote verscheidenheid aan herstelprocessen, in twee verschillende chemische fabrieken. In de eerste studie werden bijna-ongevalsrapporten van één chemische fabriek gebruikt als basis voor de dataverzameling, en deze werden steeds opgevolgd met interviews met degenen die betrokken waren bij deze gebeurtenissen. In de tweede studie, die in dezelfde fabriek werd uitgevoerd, werd een dagboekmethode gebruikt om data specifiek over het herstellen van zelfgemaakte fouten te verzamelen, omdat de data die in de eerste studie verzameld was geen voorbeelden van zulke gebeurtenissen bevatte. In de derde studie, die uitgevoerd werd in een andere chemische fabriek, werden interviews gehouden, waarin mensen gevraagd werden om recente situaties die ze zich konden herinneren te beschrijven waarbij ze er met meer of minder succes in geslaagd waren negatieve gevolgen te voorkomen door het nemen van herstelmaatregelen. De verzamelde data werden gecodeerd om verdere analyse van en vergelijking tussen herstelprocessen mogelijk te maken. In totaal werd data verzameld over 122 herstelprocessen.

De eerste drie studies toonden aan dat de herstelprocessen altijd begonnen met de detectie van een afwijking, per definitie, en dat daarna allerlei aantallen en combinaties van stappen gericht op verklaring en tegenmaatregelen konden volgen, variërend van geen enkele dergelijke stappen tot meerdere herhalingen van beide soorten stappen, meestal eindigend met tegenmaatregelen. Verschillende subdoelen konden onderscheiden worden in de verklaringfase (probleemdefinitie en vaststellen van oorzaken) en de tegenmaatregelenfase (stabilisatie, inperking, tijdelijke correctie, en permanente correctie). De studies lieten ook zien dat het creatieve denkvermogen van mensen en het toepassen van op kennis gebaseerd gedrag heel belangrijk was voor herstelprocessen. Het is daarom belangrijk dat organisaties voldoende investeren in de proces- en fabriekskennis van hun medewerkers, om ze voor te bereiden op zelfs het onverwachte. Vooraf plannen gebeurde vooral met betrekking tot de tegenmaatregelenfase en bleek deze stappen aanzienlijk te versnellen. Daarom zouden organisaties voor situaties waar snel herstel belangrijk is de bijpassende tegenmaatregelen vooraf moeten plannen. De stappen aan het begin van herstelproces bleken het snelst voltooid te worden; het doel van deze stappen was veelal het terug onder controle krijgen van de situatie. De totale hoeveelheid tijd tussen het begin en de afronding van een herstelproces hing niet alleen af van de hoeveelheid tijd die nodig was voor elk van de herstelprocesstappen, maar ook van de tijd die verstreek tussen de stappen. Daarom zouden organisaties er altijd naar moeten streven de mogelijke gevaren die als gevolg van een afwijking kunnen optreden zo snel mogelijk weer onder controle te krijgen, vroeg in het herstelproces, voordat men toestaat dat andere taken de voortgang van het herstelproces vertragen. De aanwezigheid van andere urgente en belangrijke taken bleek het stellen van prioriteiten te beïnvloeden; deze taken leidden tot een toename in de totale hoeveelheid tijd die verstreek tussen begin en voltooiing van een herstelproces. Om zeker te stellen dat hun medewerkers de juiste beslissingen nemen in het vaststellen van prioriteiten voor taken, zouden organisaties ervoor moeten zorgen dat zij de kennis hebben die nodig is om de juiste inschattingen te maken van de urgentie en belangrijkheid van zowel herstel- als normale taken. Het betrokken zijn van anderen bleek niet alleen belangrijk te zijn voor de detectie van afwijkingen, maar ook voor latere stappen in het herstelproces, vanwege hun kwalificaties, ervaring, gereedschap dat zij hadden, of de extra capaciteit die ze konden leveren, of vanwege een combinatie van deze redenen. Daarom zouden organisaties ervoor moeten zorgen dat toegang tot en communicatie met zulke anderen zo eenvoudig mogelijk gemaakt wordt. Wat betreft het verband tussen de gebeurtenissen die aan een herstelproces voorafgaan, het proces zelf, en de eindresultaten ervan, bleken er een aantal typische patronen uit de verzamelde data. Allereerst, de aanwezigheid van technische en organisatorische fouten vergrootte de noodzaak van stappen gericht op een verklaring. De aanwezigheid van organisatorische fouten vertraagde ook het moment van detectie van de afwijking. Bovendien leidde latere detectie tot langere herstelprocessen waarin meer mensen betrokken waren; zulke processen bestonden ook uit meer stappen, en als gevolg van dit alles bleven er meer of ernstigere gevolgen over na afloop van het herstelproces. Daarom zouden organisaties zich moeten richten op het zoveel mogelijk reduceren van de tijd die verstrijkt voordat een afwijking ontdekt wordt. Tenslotte bleken gebeurtenissen waarbij technische en/of organisatorische fouten en/of meer of ernstigere potentiële gevolgen een rol speelden te leiden tot langere herstelprocessen bestaande uit meer stappen en waarbij meer mensen betrokken waren, dan gebeurtenissen waarbij alleen menselijke fouten en/of weinig of geringe mogelijke gevolgen een rol speelden, en als gevolg van dit alles bleven er meer of ernstigere gevolgen over na afloop van het herstelproces. Dit gaf aan dat er ruimte zou kunnen zijn voor verbeteringen in het vermogen te herstellen van combinaties van fouten waarbij technische en organisatorische fouten een rol spelen. Organisaties zouden er echter ook voor moeten waken de huidige expertise met betrekking tot het herstellen van afwijkingen waarbij alleen menselijke fouten een rol speelden niet kwijt te raken. De mogelijke gevolgen van zulke fouten als geen herstelstappen ondernomen worden kunnen ook zeer ernstig zijn.

Twee vervolgstudies werden uitgevoerd onder operators in twee andere chemische fabrieken, om een vollediger inzicht te krijgen in de factoren die invloed hebben op het stellen van prioriteiten en het verdelen van taken onder de medewerkers in situaties waarin herstel nodig is. In tegenstelling tot de eerdere studies, werd in deze vervolgstudies rekening gehouden met de combinatie waarin deze invloeden op prioriteitsstelling en taakverdeling aanwezig waren. Bovendien, terwijl in de eerdere studies voor elk herstelproces alleen gegevens beschikbaar waren met betrekking tot de door de betrokken personen genomen stappen, werd in de vervolgstudies hetzelfde scenario van een situatie waarin herstel nodig was aan verschillende deelnemers voorgelegd. De voor de studies ontwikkelde scenario's, ofwel vignettes, bevatten elk zowel een abstracte beschrijving van een situatie waarin herstel nodig is als een voorbeeld. De waarden of niveaus van de factoren waarvan verwacht werd dat ze invloed hebben op prioriteitsstelling en taakverdeling varieerden tussen deze scenario's. Aan elke deelnemer werd een deel van de gehele verzameling van vignettes gegeven. Ze moesten voor elk scenario aangeven wanneer de nodige hersteltaken zouden beginnen, en in hoeverre ze anderen zouden betrekken in de hersteltaken. Omdat hierbij niet alleen de in de scenario's ingebouwde factoren, maar ook factoren op persoonsniveau een rol zouden kunnen spelen, werd aan alle deelnemers ook een vragenlijst gegeven om hun aantallen jaren relevante werkervaring vast te stellen, en hun vertrouwen in eigen kunnen op het werk. De met de vignettes en vragenlijsten verzamelde data werd geanalyseerd met multilevel analyse en modelbouw technieken. Door deze technieken te gebruiken werd er rekening gehouden met het hiërarchische of geneste karakter van de data (vignettes genest in deelnemers). In totaal namen 222 operators deel aan de vervolgstudies, die samen 3540 scenario's ingevuld terug inleverden.

De vervolgstudies toonden aan dat urgentie en belangrijkheid invloedrijke factoren waren bij het stellen van prioriteiten in situaties waarin een afwijking hersteld diende te worden, hetgeen resultaten bevestigde afkomstig uit time management onderzoek, en theorieën uit de psychologie met betrekking tot actie, motivatie en doelen. Dit resultaat onderstreepte nogmaals het belang voor organisaties om ervoor te zorgen dat hun medewerkers de juiste inschattingen maken van de urgentie en belangrijkheid van zowel herstel- als normale taken. De beschikbaarheid van anderen bleek een belangrijke factor te zijn voor taakverdeling, hetgeen opnieuw het belang aantoonde van toegang tot en communicatie met zulke anderen, in situaties waarin hun assistentie nodig is. De aard van de andere taken waarmee mensen bezig waren op het moment waarop ze een afwijking ontdekten, dat wil zeggen hoe belangrijk en urgent deze taken waren, speelde ook een belangrijke rol in de taakverdeling, en zo ook de beschikbaarheid van de nodige kennis en expertise. De vervolgstudies toonden aan dat medewerkers onder bepaalde omstandigheden geneigd waren af te wijken van wat de normale werkwijze of procedures voorschreven met betrekking tot wie in de hersteltaken betrokken dienden te worden. Voornamelijk de beschikbaarheid van anderen, andere taken, en kennis en expertise bepaalden de waarschijnlijkheid van zulke afwijkingen van de procedures. Deze procedureafwijkingen waren niet per definitie verkeerd; de bedoeling van de betrokkenen was om de klus te klaren en negatieve gevolgen te voorkomen. Het is natuurlijk van cruciaal belang dat wanneer mensen afwijken van wat procedures voorschrijven, degenen die taken gaan uitvoeren die eigenlijk niet voor hen bedoeld zijn weten wat ze doen, anders ontstaan er extra risico's.

Het belang van de conclusies van de vijf empirische studies waarop dit proefschrift is gebaseerd is niet beperkt tot het domein van de chemische procesindustrie. Hoewel de exacte vormen die herstelprocessen aannemen, de beschikbare tijd, capaciteit, en dergelijke kunnen verschillen in andere domeinen, zijn de algemene procesmechanismen en bijbehorend gedrag en het nut van aandacht voor herstelprocessen, zoals beschreven in dit proefschrift, onverminderd van toepassing. Ook al heeft het in dit proefschrift beschreven onderzoek al bijgedragen aan het bestaande inzicht in herstelprocessen, toch blijven er op dit gebied nog heel wat zaken over die nog verder onderzoek behoeven. Het zou bijvoorbeeld erg interessant kunnen zijn om specifieke manieren of methoden te ontwikkelen om de herstelmogelijkheden binnen een organisatie uit te breiden om zo de prestaties op dit gebied te verbeteren, deze te implementeren, en achteraf het effect ervan te meten.

Biography

Lisette was born in Terneuzen, the Netherlands, on 3 February 1971. In 1994, she graduated with an MSc degree in Industrial Engineering and Management Science from Eindhoven University of Technology in the Netherlands. In 1995, she gained a second MSc degree in Industrial Safety Management at Central Missouri State University in the USA.

In 1996, upon her return to Europe, Lisette worked at the University of Liège in Belgium on an EC funded project aiming at the exchange of safety management and accident prevention knowledge between Eastern and Western Europe.

From August 1996 to March 2000, she worked as a consultant and researcher for Human Reliability Associates (HRA), based in Dalton, Wigan (near Manchester, UK), an international consulting company specializing in improving human performance and minimizing human error in industrial sectors such as power generation, onshore and offshore chemical processing, and transport systems.

In November 1999, Lisette joined the Human Performance Management Group at the Department of Technology Management, Eindhoven University of Technology, to start working on the research resulting in this thesis.

After defending her thesis on 7 January 2004, Lisette is planning to take up a Postdoctoral Research Fellowship at the Key Centre for Human Factors and Applied Cognitive Psychology at the University of Queensland in Brisbane, Australia.

Propositions

supplementing the PhD thesis

Recovery uncovered:
How people in the chemical process industry recover from failures

by

Lisette Kanse

7 January 2004

I. It is **not** too late to lock the stable after the horse has bolted (cf. the definition of *stabilization* given in chapter 3 of this thesis).

II. If anything can go wrong, it will (captain Edward A. Murphy, US Air force, 1949 – Murphy's law) but being prepared makes a difference in the end results (chapter 3 of this thesis).

III. Linear recovery process phase models are an overly simplified way of describing recovery processes (chapter 3 of this thesis).

IV. Organizations with near miss reporting systems only use part of the potential of these systems (chapter 3 of this thesis).

V. Believing in one's ability at work has an effect on a person's willingness to take on a recovery task outside his or her normal responsibilities only when this coincides with actually knowing how to do it (chapter 4 of this thesis).

VI. For an appropriate analysis of vignette study data, the multilevel character of the data should be taken into account (cf. Hox, J.J., Kreft, I.G.G., & Hermkens, P.L.J. (1991). The analysis of factorial surveys. *Sociological Methods & Research, 19*(4), 493-510; and chapter 4 of this thesis).

VII. The success of the data collection part of a behavioral research project depends not only on the IQ (intelligence quotient), but to a large extent also on the EQ (emotional intelligence quotient) of the researcher who collects the data.

VIII. Even organizations with a successfully established near miss reporting system should regularly check the representativeness of the reported events (cf. Van der Schaaf, T.W. & Kanse, L. (2004). Biases in incident reporting databases: An empirical study in the chemical process industry. *Safety Science, 42*(1), 57-67).

IX. With increasing numbers and severity of terrorist attacks, security and safety start to overlap more and more.

X. While copying behavior is encouraged and even commended in early childhood, from a certain age onwards the opposite often applies.

XI. Progress is not served by vision that is not acted upon, nor by actions that are not guided by vision.

XII. The Australian government's practice to put asylum seekers in detention centres may very well turn them into people who belong in detention centres.

XIII. All along, there were incidents and accidents; there were hints and allegations (from "You can call me Al" – Paul Simon, 1986, Graceland album, track 6).