



NLR-CR-2007-156

**Literature survey of safety modelling and analysis
of organizational processes**

Eurocontrol CARE Innovative Research III

S.H. Stroeve, A. Sharpanskykh and H.A.P. Blom



NLR-CR-2007-156

Literature survey of safety modelling and analysis of organizational processes



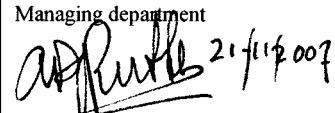
Eurocontrol CARE Innovative Research III

S.H. Stroeve, A. Sharpanskykh¹ and H.A.P. Blom

¹Vrije Universiteit Amsterdam

Customer	Eurocontrol
Contract number	C06/12396BE
Owner	Eurocontrol
Division	Air Transport Air Transport
Distribution	Unlimited
Classification of title	Unclassified
	May 2007

Approved by:

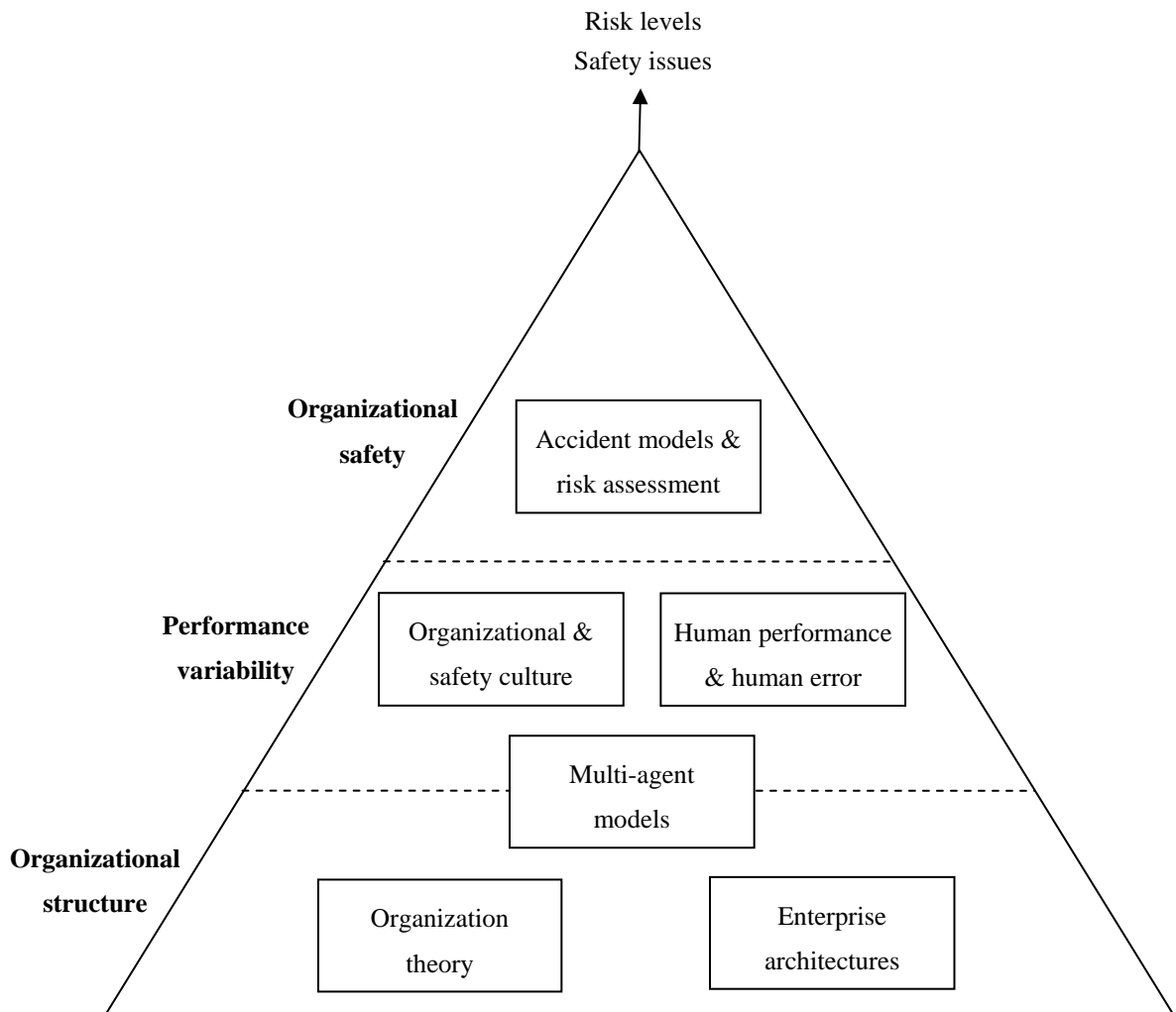
Author  20-11-07	Reviewer  21/11/2007	Managing department  21/11/2007
---	--	--

Summary

In complex and distributed organizations like the air traffic industry, safe operations are the result of interactions between many entities of various types at multiple locations. It is now generally well realized that the achieved level of safety of an operation depends on the constraints (e.g. norms, regulations, resources) set by people working at the blunt end (e.g. managers), which determine the working conditions and thereby the performance of people (e.g. pilots, controllers) who are directly controlling hazardous processes at the sharp end. In the literature and in the risk assessment practice, the recognition of the importance of organizational processes for safe operations has been accommodated by high-level conceptual models and to some extent by organizational influencing factors in sequential and epidemiological accident models. In these accident models, organizational aspects may have impact on the assessed likelihoods of causes and effects. In relation there is a considerable interest in organizational and safety culture. However, formal models that describe the variability of organizational processes and its effect on safety-relevant scenarios are largely lacking in the current risk assessment practice.

As a way forward for description of organizational processes and inclusion thereof in air traffic safety assessment methods, NLR and Vrije Universiteit Amsterdam collaborate in an Eurocontrol CARE Innovative Research III project. The objective of this research project is to enhance safety analysis of organizational processes in air traffic by development of formal approaches for modelling, simulation and analysis of organizational relationships and processes. The research includes a literature survey to identify relevant research and application of identified methods to a safety-relevant organizational process in air traffic. The current report presents the results of the literature survey.

The literature survey includes a wide variety of research disciplines, such as organizational theory, enterprise modelling, multi-agent modelling and design, organizational and safety culture, human performance, and accident risk models that cover organizational aspects. The relation of these subjects to safety assessment of organizational processes can be described at three levels in an organizational safety pyramid as depicted in the figure below: (1) organizational structure, (2) performance variability and (3) organizational safety. This three-level view fits well within the systemic accident model view, which considers accidents as emergent phenomena from the variability of a system (or organization) as a whole (Hollnagel 2004).



The first level describes the organizational structure, which encompasses human and technical organizational entities and various types of relations (e.g. power, communication) between these entities that occur in the executions of organizational work processes. The organizational structure may be specified at different aggregation levels ranging from general regulations for the whole organization to specific prescriptions for particular roles and their interaction with other roles. Analysis of processes and relations both at the same aggregation level and across different levels may show (potentially safety-critical) inconsistencies.

Multi-agent models cross the boundary between the first and second level in the organizational safety pyramid. On the one hand, their interactions are regulated by the organizational structure. On the other hand, the dynamic and stochastic aspects of interacting agents contribute to the performance variability in an organization.

The second level describes key aspects of sources for performance variability in an organization. As prime source of performance variability it addresses human performance.



Human performance is flexible and adaptable: work may be done faster/slower, tasks may be done more or less precisely, external cues may be detected or not, tasks may be omitted or done in various orders, etc. Within a systemic accident model this variability should be described to an extent appropriate for the role of the human operator within the blunt side or sharp side in the organization. The organizational or safety culture can be seen at the second level as an important moderator of the variability in the human performance. For instance, the organizational culture may be such that particular control checks are frequently omitted or that high work loads are considered normal. In this view, organizational culture (including norms, values, etc.) is made specific by indicating its effect on the variability in human performance.

The third level describes how the variability in performance of humans, the variability in the performance of technical systems, and the variability in the performance of interacting humans, interacting humans and technical systems, and interacting technical systems may lead to inconsistencies, problems, incidents and accidents. Thus, at this level accident causation may actually be described as resulting from the dynamic and stochastic interactions between the agents in the organization.

Key within the presented organizational safety pyramid is the local perspectives and behaviour of separate actors in organizational structures. Multi-agent models provide a suitable framework to represent these actors and their interactions. Accidents are considered as emergent phenomena from the variability of the agents' performance in the organizational context. In multi-agent systems conflicts can be discerned at various levels: (1) at a physical level, e.g. multiple aircraft striving to use a runway as soon as possible, (2) at a knowledge level, e.g. pilots and controllers using different information, or (3) at an organizational level, e.g. organizational norms and prescriptions imposed on the agents may be in conflict with their mental attitudes (goals, wishes, desires etc.). For the safety of complex air traffic organizations, understanding of the emerging of conflicts between its constituent agents is of key interest.

Contents

1	Introduction	7
2	Literature review	9
2.1	Overview	9
2.2	Organization theory	10
2.3	Enterprise architectures	13
2.4	Agent-based modelling	18
2.5	Multi-agent design methodologies and applications	27
2.6	Organizational and safety culture	36
2.7	Human performance and human error	41
2.8	Accident models incorporating organizational aspects	45
3	Evaluation of potential application cases	51
4	Concluding remarks	55
	References	57
Appendix A	Summaries of risk assessment methods	69

1 Introduction

In complex and distributed organizations like the air traffic industry, safe operations are the result of interactions between many entities of various types at multiple locations. Such organizations can be described at various aggregation levels. At a high aggregation level such a description discerns companies/corporations (e.g. air traffic control centres, airlines, airports, regulators), zooming in at lower aggregation levels it discerns roles and responsibilities of departments/groups, and at the lowest aggregation level it distinguishes the performance of single human operators in their organizational habitats, usually including knowledge and procedure intensive interactions with technical systems and other human operators (e.g. pilots, air traffic controllers, maintenance personnel, supervisors). In safety-focused organizations like airlines and air traffic control centres, it is important to have a good understanding of the roles and responsibilities of agents in the organization and of the interactions between the agents. Misconceptions or inconsistencies about agents' roles and responsibilities, or communication / coordination problems between agents may enhance the likelihood of development of incidents and accidents. From a safety point of view, it is important to understand these roles and interactions of agents in the organization at different aggregation levels of the organization.

The importance of proper organizational processes for the safety of complex operations is currently well realised, not in the last place due the contribution of Reason (1997). It is now generally acknowledged that the level of safety achieved in an organization depends on the constraints and resources set by people working at the blunt end (e.g. managers, regulators), which determine the working conditions of practitioners who are directly controlling hazardous processes at the sharp end.

In the literature and in the risk assessment practice, the recognition of the importance of organizational processes for safe operations has been accommodated in air traffic risk assessment by high-level conceptual models and to some extent by organizational influencing factors in accident models. Predominantly, formal risk assessment approaches focus on fault/event tree type of analysis of causes and consequences of identified hazards and organizational aspects may have impact on the assessed likelihoods of causes and effects.

Recent views on accident causation indicate that these types of accident models may not be adequate to represent the complexity of modern socio-technical systems (Hollnagel 2004, Leveson 2004, Sträter 2005, Hollnagel et al. 2006). Determinants of this complexity include the number and variety of organisational entities (human, groups, technical systems), the number and types of interdependencies between organizational entities, the degree of distribution of the entities (single/multiple locations), the types of dynamic performance of the entities (static/slow/fast), and the number and types of hazards in the organisation. Limitations of frequently applied accident models as fault/event trees include the difficultness to represent the large number of interdependencies between organisational entities and the dynamics of these

interdependencies. Since the focus on analysis of the effects of failures in these accident models is also used for the evaluation of human performance, the roles of humans in such analyses are practically restricted to making errors and resolving safety-critical situations.

To adequately account for the effects of the complexity of socio-technical organizations in safety assessment, recent views indicate that we need analysis approaches that account for the variability in the performance of interacting organizational entities and the emergence of safety occurrences from this variability. In the terminology of Hollnagel (2004) this is a systemic accident model. The systemic view considers accidents as emergent phenomena from the variability of an organization and thus passes the limitations of sequential accident models in accounting for the dynamic and non-linear nature of the interactions that lead to accidents. In the current air traffic risk assessment practice, formal models that describe the variability of organizational processes and its effect on safety-relevant scenarios are largely lacking.

As a way forward for description of organizational processes and inclusion thereof in air traffic safety assessment methods, NLR and Vrije Universiteit Amsterdam collaborate in an Eurocontrol CARE Innovative Research III project. It is the objective of this research project to enhance safety analysis of organizational processes in air traffic by development of formal approaches for modelling, simulation and analysis of organizational relationships and processes. These models must be able to describe the organization at different aggregation levels and should lead to the emerging of safety issues due the performance variability and interactions of organizational entities. It is thus intended to develop a systemic accident model for air traffic organizational issues. This development of enhanced formal methods for organizational safety modelling analysis supports a core objective of air traffic management: ensuring safe air traffic.

As a first step towards realising these objectives, a literature survey on organizational safety modelling and analysis has been performed from a wide variety of research disciplines, such as organizational theory, enterprise modelling, multi-agent modelling and design, organizational and safety culture, human performance and accident risk models that cover organizational aspects. The current report presents the results of this literature survey on modelling of organizational processes and methods for organizational safety assessment, as well as an evaluation of air traffic application cases as basis for the selection of an appropriate application case for this research project. Section 2 presents a review of the literature. Section 3 describes the evaluation of potential application cases. Section 4 provides concluding remarks.

2 Literature review

2.1 Overview

The literature reviewed for safety modelling and analysis of organizational processes considers a wide variety of sources and viewpoints:

- Organization theory, describing views on structures and dynamics of human organizations (Section 2.2) ;
- Enterprise architectures, describing models of business processes, information systems and personnel (Section 2.3);
- Multi-agent models, describing models of interacting agents for the representation of complex systems and the behaviour emerging from the multi-agent interactions (Sections 2.4 and 2.5);
- Organizational and safety culture, describing culture in an organization and its effect on safety (Section 2.6);
- Human performance and human error, describing of human performance in an operational context and the effect of its variability on the evolution of safety-relevant events (Section 2.7);
- Accident models, describing views and models for accident causation in an organizational context (Section 2.8).

The relevance of these subjects for safety assessment of organizational processes can be illustrated by the organizational safety pyramid shown in Figure 1. It describes the identification of safety issues and the evaluation of risk levels for organizational processes at three levels. The first level describes the organizational structure, which encompasses human and technical organizational entities and their relations. The second level describes sources of performance variability in an organization. The third level describes how the variability in the performance of the organizational entities may give rise to incidents and accidents. This three-level view fits well within the systemic accident model view, which considers accidents as emergent phenomena from the variability of a system (or organization) as a whole (Hollnagel 2004).

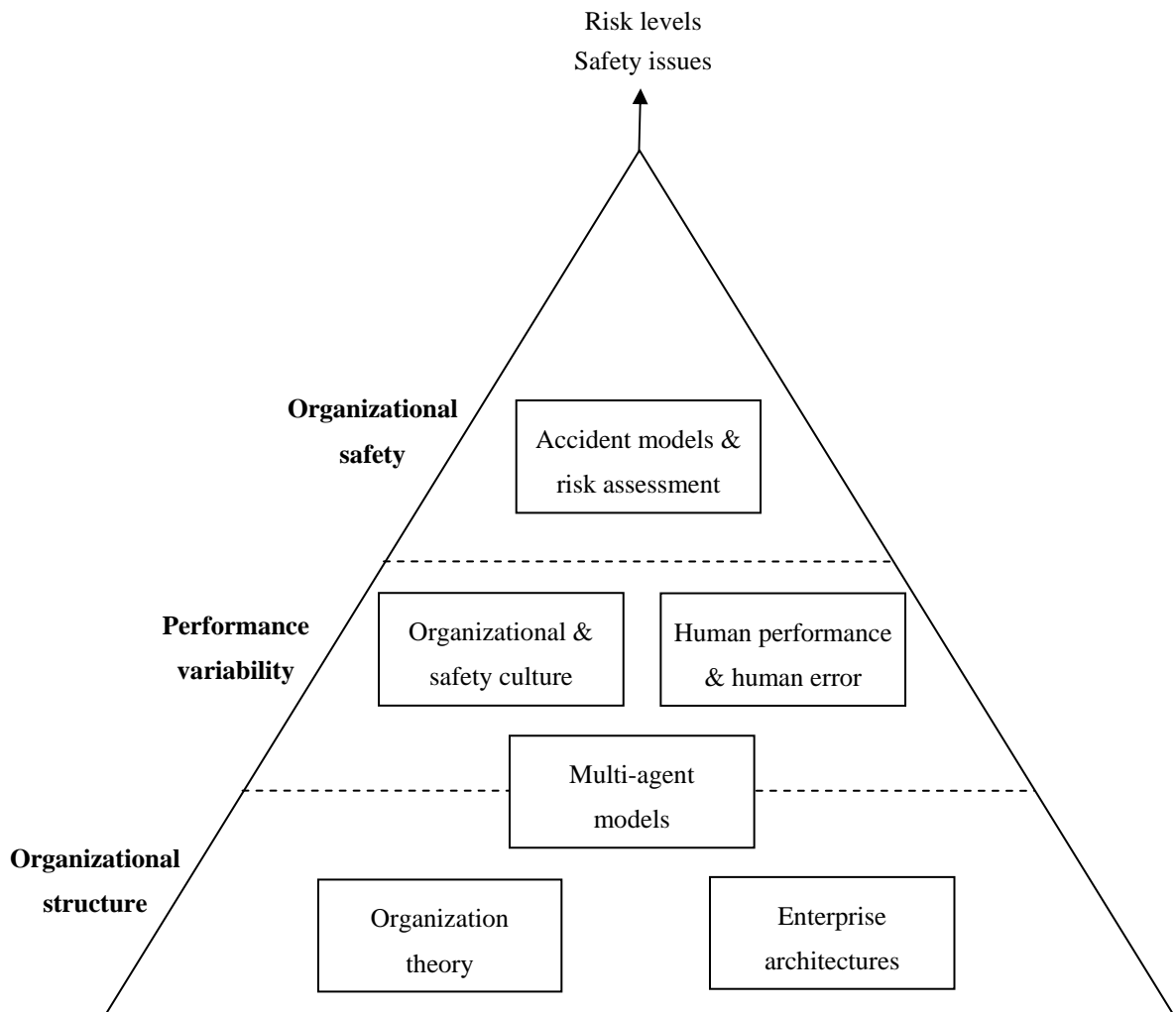


Figure 1: Organizational safety pyramid.

2.2 Organization theory

Organization theory is a broad discipline that studies structures and dynamics of human organizations. The research methods that are used in organization theory stem from such disciplines as economics, psychology, sociology, political science, anthropology, and system theory. Related practical disciplines include human resources and industrial and organizational psychology. This literature review focuses in particular on the major theories and trends in the western sociological tradition.

Definitions

Depending on the perspective, different definitions for an organization are formulated in organization theory:

- An organization is defined as a planned, coordinated and purposeful action of human beings to construct or compile a common tangible or intangible product (Giddens 2006).



- An organization is a social arrangement which pursues collective goals, which controls its own performance, and which has a boundary separating it from its environment (Scott et al. 1981).
- An organization is a structure that comprises sets of interrelated roles, which are intentionally organized to ensure a desired (or required) pattern of activities (Biddle 1979).
- An organization is defined as a system that represents an organized collection of parts that are highly integrated in order to accomplish an overall goal (Kast and Rosenzweig 1972).

The definitions by (Giddens 2006), (Scott et al. 1981) and (Biddle 1979) are formulated from the positions of sociology. The definition by (Kast and Rosenzweig 1972) is given from the perspectives of system theory.

Although the definitions given above reflect different aspects of the organizational reality, all of them are based on the concept of rationality that lies in the basis of organizational theory (Pfeffer 1982). Indeed, according to the definitions, organizations are created for certain purposes (or goals). To achieve these goals organizational activities are intentionally planned, coordinated and executed (e.g. using scientific methods).

The boundaries between an organization and its environment (e.g., other organizations, customers) may be described in many different ways and are not defined precisely in social science. However, for (formal) computational modelling and analysis of organizations, the boundaries that distinguish the modelled part of the reality (an organization) and the rest of the world should be defined explicitly and accurately. A discussion on the definition of boundaries of organizational models is given in Section 2.3.

Aggregation levels

In organization theory organizations are investigated at different aggregation levels. In particular, at the individual (sometimes called “micro”) level the behaviour of organizational individuals and groups is investigated. At the level of the whole organisation (sometimes called “meso” level) different aspects of the organisational structure and dynamics are considered. At the global (sometimes called “macro”) level the interaction between the organization and its environment that includes other organizations, society, markets etc. is considered.

Among the topics that are considered at the *individual (micro) level* are the following:

- perception of an individual in the organizational context (Scott et al., 1981);
- work motivation and satisfaction (Vroom 1964);
- the influence of personal and/or organizational values on the motivation and work-related behaviour of an individual (Yukl 2006; Hackman 1980);
- group formation (Campion, Medsker and Higgs 1993);
- group norms and regulations (Scott et al. 1981);
- social influence and conformity (Yeatts 1998);
- leadership (Yukl 2006);

- individual conflicts in organizations (March and Simon 1967);
- power and influence in groups (Yukl 2006).

At the level of the *whole organization (meso)* the following topics are of relevance:

- organization structure and behaviour (Blau and Schoenherr 1971; Mintzberg 1979; Morgan 1996);
- organization authority and power structures (Scott et al. 1981; Mintzberg 1979; Pfeffer 1982);
- organization normative systems (Scott et al. 1981);
- intergroup conflict within an organization (March and Simon 1967);
- organization reward system (Galbraith 1978; Vroom 1964);
- technology in organizations (Morgan 1996; Scott 1981);
- organizational change (Cummings and Worley 2005).

At the *global (macro) level* the behaviour of organizations is investigated using the population ecology theory (Hannan and Freeman 1977) and the resource dependence theory (Pfeffer and Salancik 1978). The following topics are considered at this level:

- inter-organizational formations (e.g., mergers and consolidations, joint ventures and programs) (Scott, Mitchell and Birnbaum 1981);
- governmental impact on organizations (Morgan 1996);
- organizations and politics (Bacharach and Lawler 1980);
- interactions between organizations and the society (Scott, Mitchell and Birnbaum 1981);
- organizations and markets (Langlois and Robertson 1995);
- virtual organizations (Warner and Witzel 2004).

Organization types

Organization theory concern organizations of different types. Classical organization theories (Mooney 1947) provide useful insights into the functioning of mechanistic organizations. This type of organizations comprises systems of hierarchically linked job positions with clear responsibilities that use standard well-understood technology and operate in a relatively stable (possibly complex) environment.

In contrast to mechanistic (or functional) organizations, a substantial group of modern organizations are characterized by a highly dynamic, constantly changing, organic structure with non-linear behaviour. Such organizations (sometimes called organic organizations (Morgan 1996)) can be investigated using modern organization theories. Modern theories are based on two essential frameworks: the systems framework (Walter 1968) and the contingency approach (Donaldson 2001).

The systems framework is based on the notion of interdependency, which implies that a change in one part of an organization affects the behaviour of all other parts. The systems framework is applied for studying matrix and network organizations (Morgan 1996).

The contingency approach (Donaldson, 2001) focuses on external determinates of organizational structure and behaviour called contingencies. A contingency is any variable that moderates the effect of an organizational characteristic on organizational performance. The key thesis of the contingency theory is that to ensure the effectiveness and the efficiency of an organization, its structure and behaviour should be defined depending on particular environmental conditions. The contingency approach is claimed to be useful for studying organizations of most of the types and it is claimed to be particularly suitable for organization design.

Organization design is a special topic in the organization theory (Lorsch and Lawrence 1970; Galbraith 1978). Galbraith (1978) stated that 'organization design is conceived to be a decision process to bring about a coherence between the goals or purposes for which the organization exists, the patterns of division of labour and inter-unit coordination and the people who will do the work.' Further Galbraith argues that 'design is an essential process for creating organizations, which perform better than those, which arise naturally.' The ideas of Galbraith and others are used extensively in the managerial practice to (re)design efficient and effective organizations (Romme, 2003). The literature on organizational design proposes an extensive set of factors identified at every level of representation of an organization (i.e., micro, meso, and macro) that influence the choice of specific design parameters (e.g., the group size, the task complexity, reporting relations, the number of employees) related to the organizational structure and dynamics.

2.3 Enterprise architectures

Enterprise architecture is an enterprise-wide, integrating framework used to represent and to manage enterprise (business) processes, information systems and personnel, so that key (strategic) goals of the enterprise are satisfied. Typically a framework for enterprise architecture includes the following aspects:

- modelling framework that consists of a set of modelling concepts (i.e., an ontology) and of a modelling language;
- modelling methodology;
- partial models (templates).

Enterprise architectures form a basis for the development of enterprise information systems used to support enterprise management (Checkland and Holwell 1998). Such systems constantly collect data about the execution of various business processes in manufacturing and production, finance and accounting, sales and marketing, and human resources. The collected data are used for different purposes. For example, based on these data inconsistencies and variances that may

occur during the execution of processes can be identified. Furthermore, these data can be used as an input for decision making processes performed by managers. Finally, these data can be provided as input for other business processes. Next to general enterprise information systems, more specialized automated systems are also used in modern enterprises. Among them are Material Resource Planning systems (MRP), Customer Relationship Management systems (CRM) and Supply Chain Management system (Gronau 2004).

Enterprise architectures may address methodological issues related to the enterprise modelling. Such issues include the identification of aspects of the organizational reality that should be captured by an organization model. Usually the specific boundaries of a model are dependent upon the specific modelling goals. For example, for the analysis of processes in a supply chain, particular organizations that constitute the chain can be modelled as parts of the same organizational model. Furthermore, enterprise methodologies address a process of engineering of enterprise models. Enterprise model engineering may be expressed in the form of a process model or structured procedure with detailed instructions for each enterprise engineering and integration activity. Several existing enterprise architectures do not have a formal foundation (e.g., CIMOSA, ARIS) that allows formal verification and validation of enterprise models built using these architectures. Such architectures do not provide means for automated analysis of data, collected by enterprise information systems that can be useful for managers.

GERAM

In the past many different enterprise architectures have been developed (CIMOSA 1993; GRAI/GIM (Doumeingts et al. 1998); TOVE (Fox et al. 1997); ARIS (Scheer and Nuetgens 2000)). Based on common features, characteristics and elements of these architectures a generalised framework called GERAM (the Generalized Enterprise Reference Architecture and Methodology) has been developed (GERAM 2003). The GERAM framework provides a generalized template for the development of elaborated enterprise modelling frameworks, based on which several international standards for enterprise modelling have been created (CD 14258, CEN/TC310, ENV 40003 and ENV 12204). Figure 2 shows a diagram of the three dimensions of the GERAM modelling framework.

- The life-cycle phases dimension describes phases in the life cycle of an enterprise.
- The instantiation dimension describes the model instantiation ranging from generic, partial to particular.
- The views dimension describes enterprise activities from the viewpoint of content, purpose or implementation.

These dimensions are further explained below.

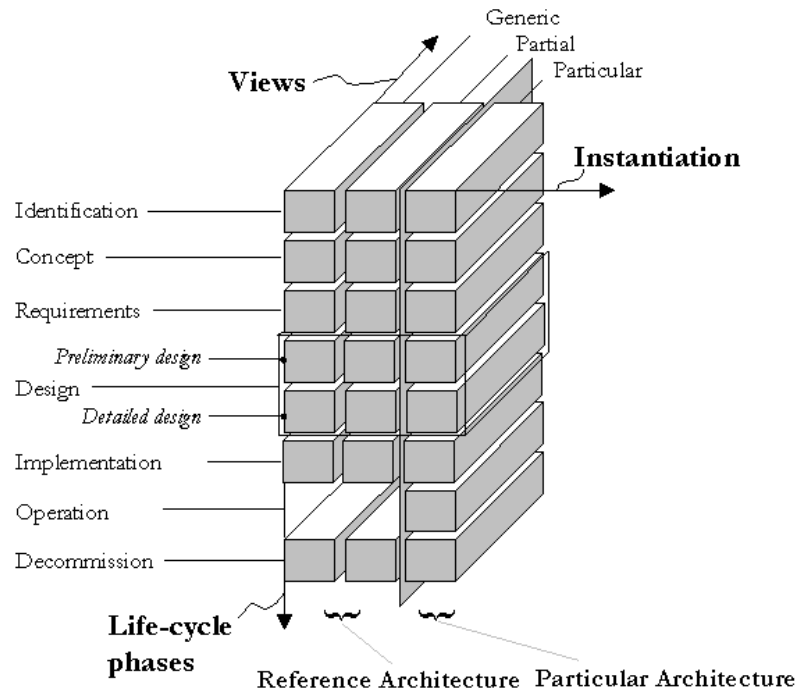


Figure 2: The GERAM Framework (adopted from (GERAM 2003)).

Life cycle phases dimension

To compare enterprise architectures GERAM identifies a number of phases of the life cycle of an enterprise that can be supported by enterprise information systems (see Figure 2):

1. Identification of the boundaries of the enterprise and the relations between internal and external environments;
2. Concept phase, at which the enterprise's mission, vision, values, strategies, objectives, operational concepts, policies and business plans are identified.
3. Requirements phase, at which operational requirements of the enterprise, its relevant processes and the collection of all their functional, behavioural, informational and capability needs are identified;
4. Preliminary design (translation of general user requirements into system requirements);
5. Detailed design (assignment to concrete human, software, hardware components);
6. Implementation;
7. Operation;
8. Decommissioning.

Instantiation dimension

An instantiation of an enterprise model can be generic, partial or particular.

- Within the *generic* dimension the most generic concepts of enterprise modelling are defined in the form of a meta-model (e.g. entity relationship meta-schema) describing the relationship among modelling concepts available in enterprise modelling languages.
- *Partial* models capture characteristics common to many enterprises within or across one or more industrial sectors. Thereby these models capitalise on previous knowledge by allowing model libraries to be developed and reused in a 'plug-and-play' manner rather than developing the models from scratch.
- *Particular* models include various designs, models prepared for analysis, executable models to support the operation of the enterprise, etc. Particular representations may consist of several models describing various aspects (or views) of the enterprise.

Views dimension

To reduce the complexity of enterprise models, GERAM proposes a number of dedicated views on enterprises that address particular aspects described along the life cycle phases introduced above. The set of the views of GERAM represents another dimension for the comparison of enterprise architectures. All views of GERAM are divided into three entity groups:

- entity model content views;
- entity purpose views;
- entity implementation views.

These views will be discussed in the following.

Entity model content views

Entity model content views are dedicated for the user oriented process representation of the enterprise and are defined as follows:

- (a) *The function view* presents the functionalities (activities) and the behaviour (flow of control) of the business processes of an enterprise. The dynamics of business processes is defined by temporal (ordering) relations, which also determine resource usage/consumption/generation schemas. This view is realized in many existing enterprise architectures and methodologies. Dynamic aspects of the execution of processes are represented using the following formalisms and frameworks: IDEF standards (Menzel and Mayer 1998), statecharts, Petri-nets (Van der Aalst and Van Hee 2002), event algebra (Singh 1996), semi-formal languages such as BPML.
- (b) *The information view* describes knowledge about objects (material and information) as they are used and produced. Information is represented in the existing architectures by data models. These models comprise entities (objects), attributes (properties of entities), attribute domains, relations among entities, key constraints (expressed as formulae in the first order predicate logic), cardinalities of relations. Different data structures adopted from computer sciences and mathematics are used in the existing architectures (Schenk and Wilson 1994):

e.g., Entity-Relationship-diagrams used in databases, object-oriented representations, UML class diagrams.

- (c) *The resource view* considers resources of an enterprise. Resources are often modelled as separate entities in the existing frameworks, however, with varying level of details. For example, in (Popova and Sharpanskykh 2007a) resources are characterized by type, amount, expiration date, category (discrete or continuous), measurement units.
- (d) *The organization view* defines responsibilities and authorities on processes, information and resources. Furthermore, the representation of organizational structure that consists of roles and relations between them (e.g., authority, interaction) is considered in this view. This view is only rarely addressed in the existing architectures. Two of the exceptions are the methodology described in (Popova and Sharpanskykh 2007a) and CIMOSA.

Entity purpose views

Entity purpose views allow representing the model contents according to the purpose of the enterprise:

- (a) *The customer service and product view* addresses the mission of the enterprise entity being studied;
- (b) *The management and control view*.

Within these views (strategic, tactical, and operational) goals of an enterprise are defined, with which the business processes of the enterprise should be aligned. Furthermore, decision making activities are addressed in these views.

Entity implementation views

Entity implementation views describe implementation aspects based on the division between human- and automated tasks:

- (a) *The human activities view* represents all information related to the tasks to be done by humans. The view distinguishes between the tasks that may be done by humans (extent of humanisability) and those that will be done by humans (extent of automation).
- (b) *The automated activities view* presents all the tasks to be done by machines. This includes information related to those tasks to be carried out by mission support technology and those carried out by management and control technology (i.e. "technology tasks"). The implementation view distinguishes between the tasks which may be done by machines (extent of automatability) and those which will be done by machines (extent of automation).

2.4 Agent-based modelling

In this section the main principles and the primary aspects of agent-based modelling are discussed. Next, Section 2.5 presents design methodologies and applications of multi-agent systems.

Modern society is characterized by a high complexity and change. A high complexity of the social dynamics results from a large number of diverse local interactions among humans. Humans interact at diverse levels in different socio-technical contexts (e.g., teams, groups, organizations etc.). Traditionally, the interaction among humans has been modelled by abstracting from single interaction processes and by taking an aggregate view on the social dynamics. The relatively new agent-based modelling approaches to complex systems take into account the local perspective of a possibly large number of separate agents and their specific behaviours (i.e., interactions) in a system. Then, the global behaviour of the system *emerges* from the local distributed interactions among agents that form a part of this system. The concept of an agent is used to model both humans as well as hardware and software components of socio-technical systems.

In (Demazeau 1995) the following components of agent-based models are distinguished: *agents*, *environments*, *interactions* and *organisations*. In the following each of these components will be addressed in detail.

Definitions of agents

A number of definitions for the concept of an agent exist, including:

- An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors (Russell and Norvig 1995).
- An agent is an active object with the ability to perceive, reason, and act (Huhns and Stephens 1999).
- An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future (Franklin and Graesser 1996).

In general, most of the definitions agree that an agent is an entity that is able to perceive its environment and to act upon this environment. In the area of multi-agent systems the agent environment is understood as the agent surroundings that include both passive and active entities (e.g., other agents) (see Figure 3).

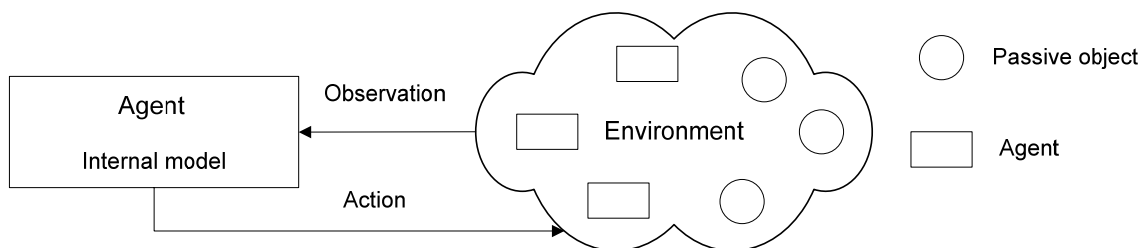


Figure 3: The classical model of an agent situated in the environment.

In most of the modern agent-based modelling approaches, agents are considered as *rational* and *autonomous* entities.

As noted in (Russell and Norvig 1995) a rational agent has the following characteristics:

- a performance measure that defines degree of success of the agent;
- a complete perceptual history (percept sequence);
- an internal representation of the environment;
- the actions that the agent can perform.

For each possible percept sequence, a rational agent should do whatever action is expected to maximize its performance measure, on the basis of the evidence provided by the percept sequence and whatever built-in knowledge the agent has.

As agent-based systems are often used for solving distributed computational problems, the rationality of behaviour of individual agents is important to ensure high efficiency and effectiveness of computational algorithms. However, for particular purposes different types of irrationalities can be specified in behavioural specifications of agents (e.g., to model human behaviour).

An agent is autonomous to the extent that its behaviour is determined by its own experience. Therefore, autonomous rational agents should be provided means to store the information received from the environment, to gain experience (or to learn) based on this information and to reason (e.g., by applying methods for problem solving, planning and decision making).

Agent types

In multi-agent systems literature, different types of agents are described with different degrees of rationality and autonomy. The notion of a bounded rational agent also appears in social science (Simon 1957), where it is defined as an agent with limited abilities in formulating and solving complex problems and in processing (receiving, storing, retrieving, transmitting) information.

In (Russell and Norvig 1995) four types of agents are distinguished:

1. *Simple reflex agents*: they do not have a mental model of the environment, and react to the stimuli from the environment in a straightforward way based on condition-action rules;



2. *Agents that keep track of the world*: reflex agents with internal states, which (re)act by executing a condition-action rule whose condition matches the information received from the environment;
3. *Goal-based agents*: goals are explicitly defined, and an agent chooses an action to achieve a goal based on the combined information about the environment, about the goal and about the results of possible actions.
4. *Utility-based agents*: a numerical performance measure (a utility) is associated with every possible (sequence of) action(s), according to the degree of its (their) contribution to the satisfaction of an agent's (high priority) goals; agents choose the action(s) with the highest utility. Behavioural specifications of utility-based agents are based on the well-known expected utility hypothesis from economics, which considers the expected utility as an expectation in terms of probability theory.

Mental models

Agent types 2 – 4 assume some kind of a mental model of an agent that includes a number of mental attitudes. A number of frameworks for the representation of mental models of agents have received a special attention in the area of multi-agent systems. Among them are Belief-Desire-Intention (BDI) and KARO frameworks. The BDI framework (Wooldridge 2000) is based on the model of human practical reasoning developed by Michael Bratman (1999) as a way of explaining future-directed intention. The perceptions of agents received from the environment are represented by beliefs in the BDI framework. Desires represent objectives or situations that the agent would like to accomplish or bring about. Intentions represent the deliberative state of the agent. Intentions are desires to which the agent has to some extent committed. The KARO framework (Van der Hoek et al., 1998) is a variant of BDI logic, which refines the motivational attitudes of agents (such as wishes, intentions, goals) described in BDI and establishes clear dynamic relations between them. Furthermore, KARO defines motivation states of an agent as a driving force for making commitments and performing actions by an agent.

Environments of agents

The agents are situated in different types of environments, from which they receive information and in which they act. In (Russell and Norvig 1995) environments are classified along the following dimensions:

1. *Accessible versus inaccessible*: if an agent's sensory apparatus gives it access to the complete state of the environment, then the environment is accessible for the agent.
2. *Deterministic versus nondeterministic*: if the next state of the environment is completely determined by the current state and the actions selected by the agents, then the environment is deterministic.

3. *Episodic versus nonepisodic environment*: in an episodic environment the agent's experience is divided into episodes. Each episode consists of the agent perceiving and then acting.
4. *Static versus dynamic*: if the environment can change while an agent is deliberating, then the environment is dynamic for this agent; otherwise it is static.
5. *Discrete versus continuous*: if there are a limited number of distinct, clearly defined percepts and actions, then the environment is discrete.

In (Weiss 1999) the environments are characterized along the dimensions given in Table 1.

Table 1. Characteristics of the environments given in (Weiss 1999).

Attribute	Explanation	Range
Predictability	The amount of uncertainty about the static and dynamic properties of the environmental objects	Foreseeable... Unforeseeable
Accessibility and knowability	The degree of accessibility of information from the environment to an agent The amount of knowledge about the environment that an agent can potentially possess	Unlimited...Limited
Dynamics	The degree of prescription of certain scenarios of the environmental behaviour	Fixed...Variable
Diversity	The amount of different types of (active and passive) environmental objects	Poor...Rich
Availability of resources	Types of access regulations for resources	Restricted...Ample

Interactions between agents

The agents situated in the environment may interact with other agents (humans, computers, software) or with passive environmental objects (e.g., databases, tools). Interaction can take place indirectly through the environment in which the agents are embedded (e.g., by observing one another or by carrying out an action that modifies the environmental state) or directly through a shared language (e.g., by providing information to agents).

In the area of multi-agent systems the focus is on coordination as a form of interaction that is particularly important with respect to goal attainment and task completion. The purpose of coordination is to achieve or avoid states of affairs that are considered as, respectively, desirable or undesirable by one or several agents (Weiss 1999). Two basic contrasting patterns of coordination are *cooperation* and *competition*.

In the case of cooperation, several agents work together and make use of their combined knowledge and capabilities to achieve a common goal. Typically, to cooperate successfully, each agent must maintain a model of other agents, and also develop a model of future interactions. Furthermore, for successful cooperation it is important to ensure that all cooperating agents create and maintain consistent, valid and precise internal representations

(e.g., belief sets) about what should be achieved (i.e., goals) and about the achievement means (Gurnell, 2006).

To enable meaningful cooperation of agents, a proper protocol should be defined. In the simplest case, such a protocol is based on the decomposition and the subsequent distribution of tasks. However, for a more effective and efficient cooperation more sophisticated cooperation strategies are defined, based on the following mechanisms (Huhns and Stephens 1999):

1. market mechanisms: tasks are matched to agents by generalized agreement or mutual selection;
2. contract net: announce, bid, and award cycles;
3. multi-agent planning: planning agents have the responsibility for task assignment;
4. organizational structure: agents have fixed responsibilities for particular tasks.

In contrast to cooperation, in the case of competition several agents work against each other because their goals are usually conflicting. Competitive agents try to maximize their benefit at the expense of others, and so the success of one implies the failure of others. The competition form of coordination is usually based on negotiation mechanisms. In (Huhns and Stephens 1999) the negotiation is defined as ‘a process by which a joint decision is reached by two or more agents, each trying to reach an individual goal or objective.’

In the area of multi-agent systems a large variety of methods for negotiations have been developed (Sycara 1998) that can be divided into two large groups: (1) environment-centred and (2) agent-centred. Developers of environment-centred techniques focus on the following problem: “how can the rules of the environment be designed so that the agents in it, regardless of their origin, capabilities, or intentions, will interact productively and fairly (i.e., according to some rules)?” Developers of agent-centred negotiation mechanisms focus on the other problem: “Given an environment in which the agent operates, what is the best strategy for it to follow?”

In (Weiss 1999) the interactions among the agents are characterized along the dimensions described in Table 2.

Table 2. Characteristics of the interactions among the agents given in (Weiss 1999)

Attribute	Explanation	Range
Frequency	Frequency of interactions among agents	Low...High
Persistency	Persistency of particular interaction patterns of agents	Short-term...Long-term
Level	Type of information exchanged between agents (i.e., the ontological expressivity of information)	Signal-passing... Knowledge-intensive
Pattern (flow of data and control)	Type of organizational structure imposed on the interaction among agents	Decentralized... Hierarchical
Purpose	The purpose of interaction between agents	Competitive... Unforeseeable



Note that the set of goals of an agent may contain both (shared) goals that conform to goals of other agents and goals that conflict with goals of other agents. In such a case the agent may be competing for the achievement of the conflicting goals, and at the same time collaborating with other agents to achieve their common goals. Complete achievement of all goals of both types may not be feasible. In such situations the agent may assign different degrees of priority and acceptable degrees of satisfaction for each goal. For example, in the context of air traffic control, an airline carrier may have the competitive goal to decrease the waiting time at the airport and the cooperative goal to maintain a high level of safety of each operation, the fulfilment of which is also dependent on other actors of an air traffic control organization. Ideally, the goals related to safety have the highest priority and should be satisfied by every actor.

Organizations of agents

Interactions among agents often take place in the context of certain organizational formations (or structures). On the one hand, such structures may be intentionally designed to enforce certain rules on the behaviour of an agent, e.g., which are based on norms and policies, organizational culture etc. On the other hand, organizational structures may ‘emerge’ from the non-random and repeated patterns of interactions among agents. The organization structure provides means to coordinate the execution of tasks in a multi-agent system and to ensure the achievement of organizational goals. Furthermore, by defining the organizational layer in agent models, the predictability of the behaviour of the agents can be substantially increased.

In (Horling and Lesser 2005) several types of organizational structures are distinguished. The most important ones are described next.

1. *Hierarchies* (see Figure 4): Agents are conceptually arranged in a treelike structure, where agents higher in the tree have a more global view than those below them. Interactions do not take place across the tree, but only between connected entities. The data produced by lower-level agents in a hierarchy typically is transmitted upwards to provide a broader view, while control flows downward as the higher level agents provide direction to those below.

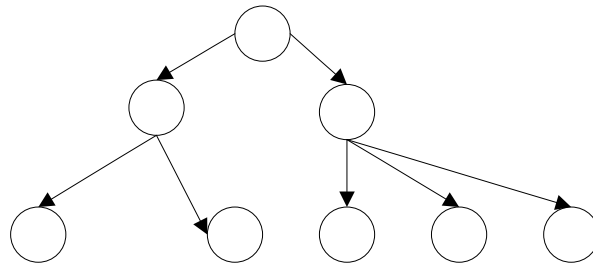


Figure 4: A hierarchical structure.

2. *Holarchies* (see Figure 5): They represent hierarchical nested structures that consist of holons. Each holon is composed of one or more subordinate entities, and can be a member of one or more superordinate holons. Many holonic structures also support connections between holons across the organization, which can result in more amorphous, weblike organizational structures that can change shape over time.

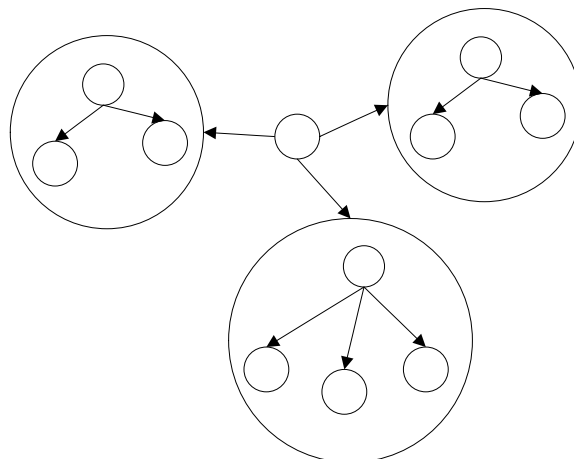


Figure 5: A holarchical structure.

3. *Coalitions* (see Figure 6): Usually flat organizations that comprise sets of agents. Coalitions in general are goal-directed and short-lived. In coalitions each agent-member pursues its own individual goal.

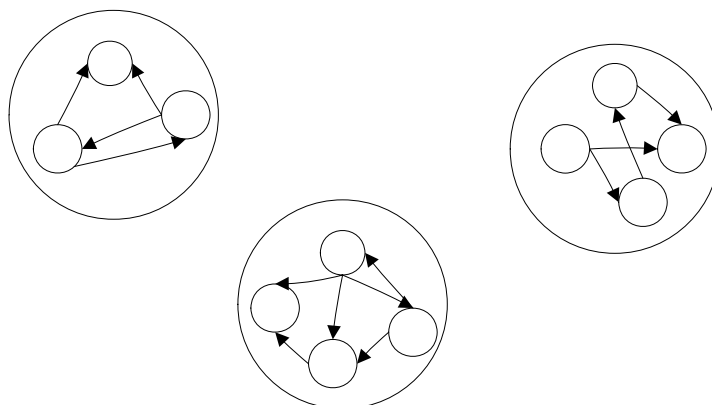


Figure 6: A coalition-based structure.

4. *Teams* (see Figure 7): An agent team consists of a number of cooperative agents which have agreed to work together toward a common goal. In comparison to coalitions, teams attempt to maximize the utility of the team (goal) itself, rather than that of the individual members. Agents are expected to coordinate in some fashion such that their individual actions are consistent with and supportive of the team's goal.

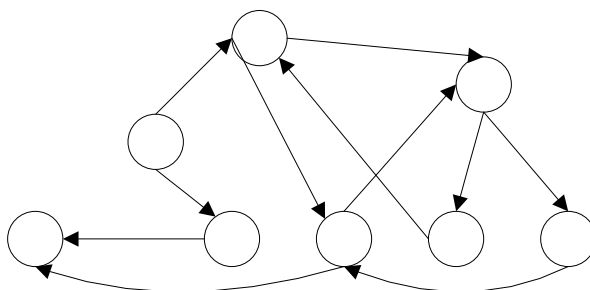


Figure 7: A team-based structure.

5. *Congregations*: Similar to coalitions and teams, agent congregations are groups of individuals who have banded together into a typically flat organization in order to derive additional benefits. Unlike these other paradigms, congregations are assumed to be long-lived and are not formed with a single specific goal in mind. Instead, congregations are formed among agents with similar or complementary characteristics to facilitate the process of finding suitable collaborators.
6. *Federations* (see Figure 8): Agent federations, or federated systems, come in many different varieties. All share the common characteristic of a group of agents which have ceded some amount of autonomy to a single delegate which represents the group. The delegate is a

distinguished agent member of the group, sometimes called a facilitator, mediator or broker. Group members interact only with this agent, which acts as an intermediary between the group and the outside world.

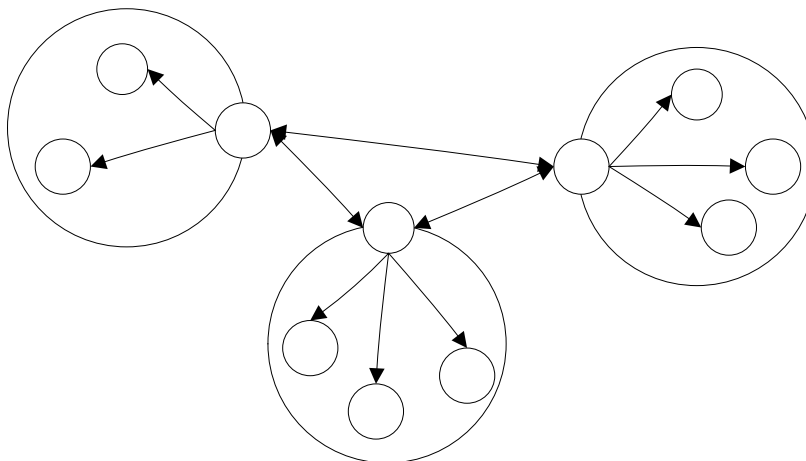


Figure 8: An agent federation.

Often organizational structures are specified in terms of roles. A *role* is usually defined as an abstract representation of a set of functionalities performed by an organization. Furthermore, roles are often characterized by sets of requirements (skills, traits and capabilities) that agents should fulfil in order to be allocated to these roles. Note that several agents can be allocated to the same organizational role, and several roles can be enacted by one agent.

Depending on the type of an organizational structure, agents are provided different degrees of autonomy. Usually the behaviour of agents is restricted by a set of norms that can be defined at different aggregation levels of the organizational structure. Currently many approaches for modelling normative multi-agent systems have been proposed in the literature. In particular, in (Grossi et al. 2006) different types of organizational norms are defined that may be specified in different types of organizations. Besides prescriptive also descriptive behavioural specifications are provided for multi-agent systems in some approaches (Jonker et al., 2006).

Many of the existing approaches describe organization models, using only two or three aggregation levels; i.e., the level of an individual role, the level of a group composed of roles, and the overall organization level. A few exceptions (e.g., Jonker et al., 2006) allow representing as many aggregation levels as required. One of the important issues considered for multi-level organization-oriented multi-agent systems is the consistency of the structural and behavioural specifications of these systems both of particular aggregation levels and across multiple levels.

Conflicts between agents

As agents represent autonomous (sometimes self-interested) active entities, different types of conflicts inevitably appear at different layers of agent-based models. In (Tessier et al. 2001) two classes of conflicts in multi-agent systems are identified: (1) physical conflicts and (2) knowledge conflicts. Physical conflicts are resource conflicts. A typical physical conflict is the intention of using at the same time a resource which cannot be shared out or which is scanty. Knowledge or viewpoint conflicts stem from the fact that the agents' information is not the same. For instance, agents may have different skills or may be equipped with different sensors. Such conflicts may concern agents' beliefs, knowledge, opinions, and are often called "epistemic" conflicts. In addition, agents may also have conflicts at the organizational layer. For example, organizational norms and prescriptions imposed on the agents may be in conflict with the specifications of mental attitudes of the agents (such as goals, wishes, desires etc.).

Two classical strategies to cope with conflicts are avoiding them and solving them. In the former case agents apply common rules and conventions (Jennings 1994) or rely on mutual representations of others' goals, intentions and capabilities (Sichman et al., 1994). More complex approaches try to represent tasks and resource dependencies (Decker and Lesser, 1992). Hence, this approach makes each agent more predictable and reduces the need for coordination. However, since the agents have limited knowledge of their environment and of the other agents, conflict solving is unavoidable in some situations. Usually, this is done by synchronization algorithms (i.e., a centralized approach (Cammarata et al., 1983) or negotiation protocols (i.e., a decentralized approach (Rosenschein and Zlotkin, 1994).

2.5 Multi-agent design methodologies and applications

This section presents subsequently design methodologies and applications of multi-agent systems.

Multi-agent design methodologies

Many different methodologies for modelling and design of multi-agent systems have been proposed. A summary of characteristics identified in the previous section for selected methodologies is given in Table 3; their details are presented next.

Table 3. Summary of characteristics for some of the methodologies for modelling and design of multi-agent systems. A '+' denotes that a characteristic is addressed in the methodology, '-' denotes that a characteristic is not consider.

Methodology	Environ- ment	Agents		Organization		Implemen- tation
		Internal models	Interaction	Structure	Dynamics	
GAIA	-	-	+	+	-	-
AGR	-	-	+	+	-	+

SODA	+	-	+	+	+	-
MOISE	-	+	+	+	-	+
TROPOS	-	+	+	+	+	+
DESIRE	+	+	+	-	-	+
(Popova and Sharpanskykh 2007e)	+	+	+	+	+	+

The GAIA methodology (Zambonelli, Jennings, and Wooldridge 2003) addresses two development phases: an analysis and design phase. The analysis phase describes two models: a role model and an interaction model (Figure 9). The role model specifies organizational roles. The interaction model defines the dependencies and relation between roles by means of protocol definitions. At the design phase societies of agents are specified. The design phase provides three models: the agent model, the service model, and the acquaintance model. The agent model identifies agent types, which are sets of roles. The service model identifies the services (or functions) associated with a role. Finally, the acquaintance model identifies the communication links between agent types. GAIA does not capture the internal aspects of agents. The interaction of agents with the environment is not treated separately.

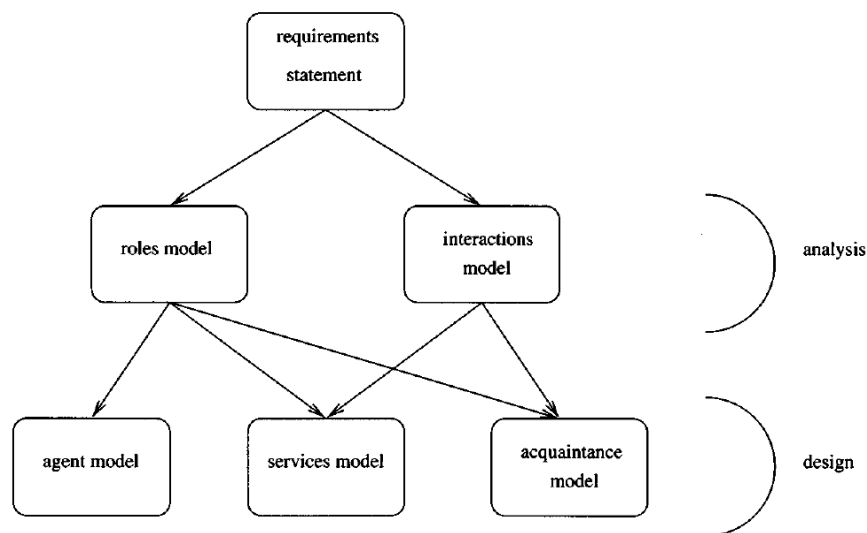


Figure 9: Relationships between GAIA's models

The AGR methodology proposed in (Ferber and Gutknecht 1998) considers both structural and dynamic aspects of organizational models. Each organization model of the AGR comprises a set of interrelated groups that consist of roles (Figure 10). Groups are related through shared agent(s) allocated to roles within these groups. Furthermore, the AGR places no constraints on the internal architecture of agents and does not provide any implementation details, except for

the recommendation to use ACL FIPA language for implementing the communication between agents. In (Ferber, Michel, Baez-Barranco 2004) an extension of the AGR is described called AGRE (AGR + Environment). This extension includes a representation of physical (or simply geometrical) environments.

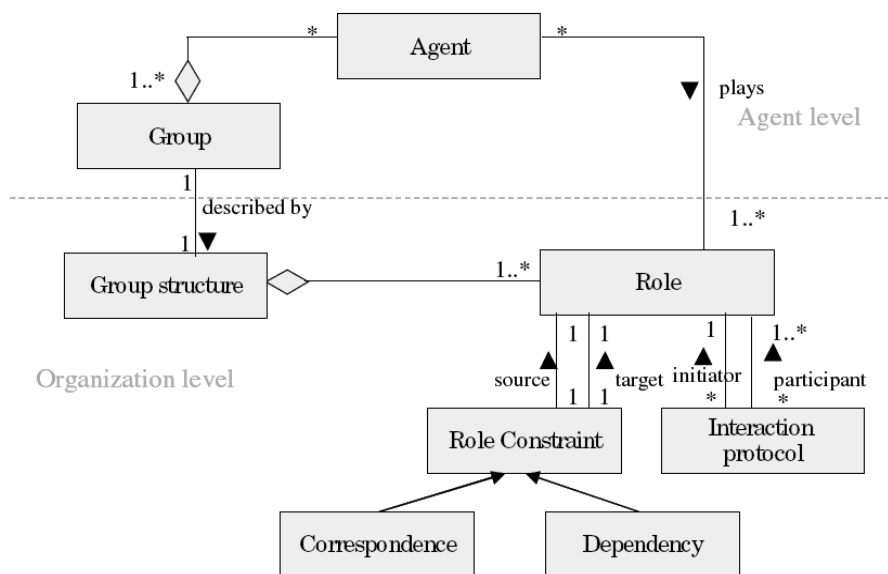


Figure 10: The meta-model of AGR

The SODA methodology (Omicini 2000) makes a distinction between the analysis and design phases of the development process. The analysis phase provides three models: the role model, the resource model, and the interaction model. The design phase refines the abstract models from the analysis phase and provides three models: the agent model, the society model and the environment model. SODA focuses particularly on inter-agent interactions and does not specify the design of the agents themselves.

The MOISE methodology (Hannoun et al. 2000) allows constructing models along three levels: (1) the individual level of agents; (2) the aggregate level of large agent structures; (3) the society level of global structuring and interconnection of the agents and structures with each other. The methodology also addresses some implementation issues of the introduced models.

The TROPOS methodology (Bresciani et al. 2004) addresses three development phases of multi-agent systems: the analysis, design and implementation phases. The analysis phase is represented by an early and a late requirements phase. The early requirements phase, which is based on the i* organizational modelling framework (Yu 1997), is concerned with understanding an application by studying its organizational setting. The late requirements phase results in a list of functional and non-functional requirements for the system. The design phase is divided into an architectural design and a detailed design phase. The architectural design

defines the structure of a system in terms of subsystems that are interconnected through data, control and other dependencies. The detailed design defines the behaviour of each component. The implementation phase maps the models from the detailed design phase into software by means of Jack Intelligent Agents (Hodgson et al. 2000).

The Design and Specification of Interacting Reasoning (DESIRE) framework (Brazier et al., 1997) is a complete environment for design and implementation of multi-agent systems. It allows the system designer to explicitly and precisely specify both the intra-agent and inter-agent functionality. In DESIRE, the following models are supported: (1) task (de-) composition, (2) information exchange, (3) sequencing of (sub-) tasks, (4) subtask delegation and (5) knowledge structure. The task (de-) composition model includes the information about the task hierarchy, the task input, the task output and the relationships between tasks. Each task in the hierarchy can be primitive as well as composed. Additional information regarding the task composition model is encapsulated within the information exchange model. In DESIRE, tasks are allocated to particular components. The information exchange between tasks is specified as information links between components. Each information link directs the output of one component to the input of another one. An information exchange model is depicted in Figure 11. Rectangles with rounded corners represent components and arrows represent information channels between components both at the same and between different aggregation levels.

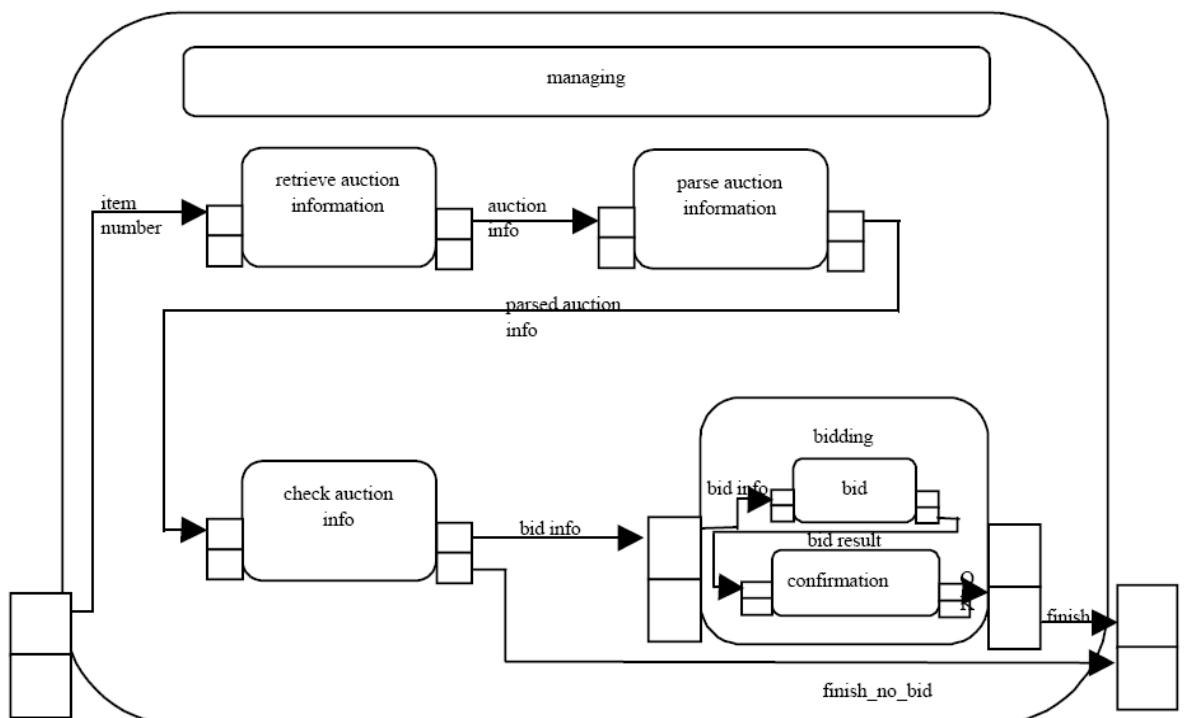


Figure 11: The component management in Desire

Task sequencing is explicitly modelled within the component as task control knowledge. This knowledge includes the order of subtasks, their activation target (usually referred to as a goal in the context of agents) and the amount of effort, which can be afforded. DESIRE has the additional advantage that the specifications and their semantics can be expressed formally, using temporal logic.

A formal framework for modelling and analysis of both human and artificial organizations based on the agent paradigm is proposed in (Popova and Sharpanskykh 2007e). A formal organizational model (Figure 12) is specified in this framework using a great variety of organizational concepts and relations that are structured into a number of dedicated perspectives (or views), similar to the ones defined in GERAM (2003): performance-oriented (Popova and Sharpanskykh 2007a, 2007b), process-oriented (Popova and Sharpanskykh 2007c, 2007d) and organization-oriented (Sharpanskykh 2007; Jonker et al. 2007). The concepts and relation of the views are formalized using the dedicated languages based on order-sorted predicate logic. The views are related to each other by means of sets of common concepts. The introduced views have a prescriptive character and define the desired behaviour of the organization. The fourth view, agent-oriented, describes and integrates agents into the framework. Models of agents are based on the extensive theoretical basis on modelling humans in organizational context developed in social science (e.g., theory of needs (Child 1973), expectancy theory (Vroom 1964)). In particular, the agent modelling concerns intentional and motivational aspects of agent behaviour. The framework proposes different types of analysis: formal verification and validation of models of different views; simulation for experimenting and testing hypothesis on the organizational behaviour under different circumstances; manifold computational analysis methods across multiple views. The framework allows the representation and analysis of organization models at different levels of abstraction in order to handle complexity and increase scalability.

The framework has been applied in a number of case studies. In particular, an organization from the security domain within the project CIM (Cybernetic Incident Management, see <http://www.almende.com/cim/>) has been modelled and analyzed for the purpose of identifying performance bottlenecks and inconsistencies in the organizational structure and behaviour (Popova and Sharpanskykh 2007a, Popova and Sharpanskykh 2007d). For this case study organizational models within different views have been created and analysed using the dedicated techniques of the corresponding views. Within the same project the approach has been applied to model and analyze the progress of incident management after the incident has occurred or while the incident unfolds (Abbink et al. 2004; Hoogendoorn et al. 2006). To this end, based on incident reports, models that reconstruct an incident management process and related structures have been specified in the form of traces (i.e., temporally ordered sequences of events). Then, structural and dynamic properties obtained from different sources (e.g., legal regulations, disaster plans, disaster prevention plans, and laws) have been checked automatically

with respect to these traces. In such a way different types of inconsistencies and errors in the actual incident management processes have been identified. By applying the proposed analysis methods, discrepancies between formal incident management prescriptions and guidelines and actual sequences of event can be identified. The framework has been also applied for modelling of organizations in the area of logistics (Jonker et al. 2007, Popova and Sharpanskykh 2007c).

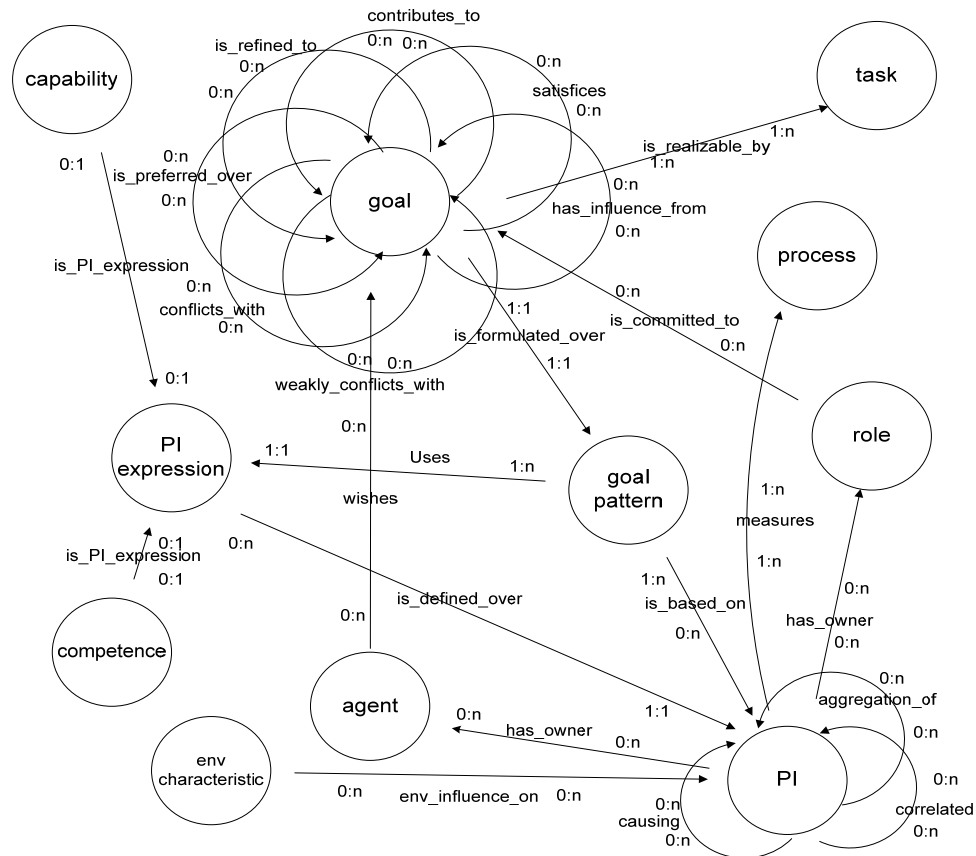


Figure 12: A part of the meta-model of the framework from (Popova and Sharpanskykh 2007a); here PI denotes Performance Indicator

A detailed comparison of three advanced methodologies TROPOS, DESIRE, (Popova and Sharpanskykh, 2007e) along the main characteristics identified in the previous section is given in Table 4.

Table 4. Summary of characteristics for three methodologies for modelling and design of multi-agent systems.

Aspect	Characteristic	TROPOS	DESIRE	(Popova and Sharpanskykh 2007e)
Environment	Determinism	Not addressed	Non-deterministic environment can be implemented	The probability of occurrence of environmental events can be defined
	Dynamics		Different behavioural scenarios for the environment can be specified	
	Diversity		The underlying language of the framework allows representing a wide variety of environmental objects and processes	
Agents	Reflex agents	Can be represented using direct stimulus-response relations		
	Agents with internal states	Represented using belief states	Represented using internal states	Represented using intrinsic memory states
	Goal-based agents	Can be specified using the goal concept and the rules for the goal generation, adoption and attainment		
	Utility-based agents	The formal framework allows only limited possibilities to reason about the agent's belief, desires and goals	Different reasoning mechanisms of agents are developed: e.g., based on motivation models, trust models etc.	
Interactions among agents	Types of interactions	Only one interaction type is addressed – communication, represented by UML activity diagrams	All types of interaction are represented by communications between components	Distinguishes communication acts and actions to represent all types of interaction
	Level	All information level types are supported		
	Purpose	Can represent all types of purposes		
Organizational model	Structures	No aggregated structures; it distinguishes the notion of a role and represents interaction relations between roles at the same aggregation level only (i.e., no means to represent hierarchical structures); no authority aspects are defined	Aggregated structures of components can be built, however different types of relations between roles that exist in real organizations (e.g., authority relations) are not distinguished	Allows representing both multi-level hierarchical structures (e.g., holarchies, federations) and flat structures (e.g., teams, coalitions); permits specifying both interaction and authority relations on roles

	Dynamics	Represents the logical dynamics of internal processes of agents by plan graphs; the dynamics of agent interaction is represented by UML activity diagrams and logic-based languages	Represented by executable <i>if-then-</i> rules. The explicit representation of time can be added, however, requires additional implementation	Represents the dynamics at a logical level by properties expressed using an expressive temporal predicate logic-based language
Implementation		Software agents are realized using the Jack Intelligent Agents platform (Hodgson et al. 2000), furthermore, dedicated verification techniques are developed	Supported by the dedicated software tool DesTool	The dedicated tools for performing agent-based simulations (Bosse et al. 2005), verification and validation of agent-based models (Bosse et al., 2005; Popova and Sharpanskykh, 2007c;d) have been implemented

Multi-agent applications

Many different cognitive agent models have been developed in the area of artificial intelligence for practical applications. For example in Sugarscape (Epstein and Axtell 1997) the agents have physical positions and follow simple rules to respond to each other and their environments; in VDT (Cohen 1992) agents are modelled as simple processors with in- and out-boxes; in CORP (Carley 1996) simple model of experiential learning is used; and in Plural-Soar (Carley et al. 1991) and TAC Air Soar (Tambe 1997) a fully articulated model of human cognition is used.

The agent coordination mechanisms described in Section 2.4 have been used for particular applications. For example, in (Davis and Smith 1983) it has been explored how the Contract Net Protocol can be used for solving a variety of problems related to the Distributed Sensor Net Establishment (DSNE).

Coordination mechanisms have been also widely applied for distributed planning. In this topic two aspects are distinguished: the formulation of a plan that can be distributed among a variety of execution systems and the planning process itself that can be distributed among agents. Distributed planning techniques have been applied in such domains as mission planning for unmanned vehicles (Durfee, Lesser and Corkill 1990) and for logistics planning (Wilkins and Meyers 1995).

Distributed decision making based on negotiation mechanisms is another important topic in the area of multi-agent systems, which includes many practical implementations. In (Pitt et al. 2006) a voting approach has been proposed that allows making decisions based on the



inputs from agents. Furthermore, the use of different types of auction protocols (e.g., English (first-price open-cry) auction, Dutch (descending) auction, Vickrey (second-price sealed-bid) auction) has been explored in many applications (Russell and Norvig 1995), such as electronic commerce and electronic markets. For exchanging information and knowledge between agents the knowledge query and manipulation language (KQML) is often used. This language underlies many existing coordination protocols.

A task environment-oriented modelling framework TAEMS for agent-based systems is proposed in (Regis, Bryan, Lesser 2000). The framework allows to analyze and to quantitatively simulate the behaviour of single or multi-agent systems with respect to characteristics of the computational task environments of which they are part. TAEMS has been applied for analyzing a simple distributed sensor network.

The framework proposed in (Hodgson 2000) allows building teams of agents. A team is a composite component, similar to a group, which is characterized by a number of roles, enacted by agents and other teams.

In the industry a number of frameworks have been developed for designing systems of aggregated (or composite) agents that are often called holons (Schillo and Spresny 2005). A holon is defined by a recursive model of agent groups and appears as a single entity to the outside world. A holon may impose certain structures (i.e., types of relations) and behaviours on its agents, thus limiting their autonomy in certain aspects. Furthermore, a holon may be allocated to a simple (not composite) role, when the joint set of capabilities of agents of the holon satisfies the role requirements.

Industrial applications of agent technology were among the first to be developed. For example, for the purposes of process control the ARCHON system has been developed (Jennings, Corera, and Laresgoiti 1995). It has been applied in several process control applications, including electricity transportation management and particle accelerator control.

Ljungberg (1992) describes an agent-realized air traffic control system OASIS. OASIS is a decision support system developed in conjunction with Airservices Australia. OASIS was designed as an intelligent assistant to the Flow Director (the person responsible for tactical air traffic management), to reduce the Flow Director's workload and allow more efficient management of arrivals. It was developed as a proof of concept prototype and successfully trialed by controllers in parallel with the existing manual approach at the Area Control Center for Sydney Airport in 1995. These trials demonstrated that a multi-agent system could work effectively in a real-time operational environment, performing many of the low-level reasoning and computational tasks previously performed manually by the Flow Director. OASIS has been implemented using the dMARS system (d'Inverno, Kinny, and Luck 1997).

An application built on the rational-agent architecture is the Procedural Reasoning System, a generic reasoning system described in (Ingrand, Georgeff, and Rao 1992). The proposed system has been applied to several real-time applications, including mobile robot

control, system control for a surveillance aircraft, and air traffic management. An application of an agent-based approach to plan, coordinate, and manage the container terminal domain is described in (Henesey, Wernstedt, and Davidsson 2003).

The importance and the applicability of the agent paradigm is increasingly recognized in the area of incident management. For example, in (Schoenharl et al. 2006) an agent-based system is introduced that provides to emergency planners and responders the possibility to detect possible emergencies, as well as to suggest and evaluate possible courses of action to deal with the emergency. Another example is an intelligent system for exploring dynamic crisis environments proposed in (Tatomir, Rothkrantz and Popa 2006). This agent-based system provides help in guiding people in damaged buildings. One more example is an extension of the standard BDI model with the capability of situation awareness for disaster relief operations management (Jakobson et al. 2006). This extension implements in MAS architecture such functions as event collection, situation identification, and situation assessment.

2.6 Organizational and safety culture

The previous sections provided views and models for the structure of organizations and the interactions between organizational entities. Such formalised policy and procedures of the organization represent one layer of an explanation of how work is actually done. It is well realized that actual work processes vary with respect to the formal rules and this variability may well be essential to get the job done in varying contextual conditions. This is a core concept in the reasoning on systemic accident models (Hollnagel, 2004). In this context, organizational culture and its subfacet safety culture can be seen as conceptual reflections on the variability in work processes. In other words, they reflect the impact on “the way we do things around here”, which is an informal, behaviour focussed notion of organizational culture (Hopkins, 2006). In this light, the current section exposes some notions on organizational and safety culture that have been addressed in the literature.

Some definitions

There are various definitions of *organizational culture*, including

- A pattern of basic assumptions – invented, discovered, or developed by a given group as it learns to cope with its problems of external adaptation and internal integration; that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems (Schein 1992);
- Shared values (what is important) and beliefs (how things work) that interact with a company’s people, organizational structures and control systems to produce behavioral norms (the way we do things around here) (Uttal 1983);



- A relatively stable, multidimensional, holistic construct shared by (groups of) organizational members that supplies a frame of reference and which gives meaning to and/or is typically revealed in certain practices (Guldenmond 2000);
- The way we do things around here (Hopkins 2006).

The definition of Schein stems from sociology, Uttal's one is from management theory, the definition of Guldenmond is based on a review of a range of definitions, and the functional one of Hopkins reflects a practitioner's view.

In addition the concept of organizational culture, some authors also use the term *organizational climate*. Guldenmond (2000) and Hopkins (2006) argue that the terms organizational culture and organizational climate have been used in various not always distinguishable ways. Currently, the term organizational climate usually indicates the overt manifestation of organizational culture. Cox and Cox (1996) compare the culture and climate of an organization with, respectively, the personality and mood of an individual.

Safety culture reflects safety-relevant aspects of organizational culture. A wide range of definitions are provided in reviews by Guldenmond (2000) and Choudhry et al. (2006), including:

- The set of beliefs, norms, attitudes, and social and technical practices that are concerned with minimising the exposure of employees, managers, customers and members of the public to conditions considered dangerous or injurious (Pidgeon 1991);
- The product of individual and group values, attitudes, perceptions, competencies, and patterns of behaviour that determine the commitment to, and the style and proficiency of the organization's health and safety management (Lee 1996);
- The attitudes, beliefs and perceptions shared by natural groups as defining norms and values, which determine how they act and react in relation to risk and risk control systems (Hale 2000).

Hopkins (2006) notes that there is confusion about the meaning of safety culture. He argues that the study of organizational culture should be distinguished from the impact of organizational culture on safety; as such there is no need to use the term safety culture.

Similarly to the current notion of organizational climate, the term *safety climate* is sometimes used to indicate the overt manifestation of safety culture.

Models of organizational/safety culture

There are several models of (aspects of) organizational and safety culture.

Regarding layers or levels of organisational/safety culture, Glendon and Stanton (2000) and Guldenmond (2000) describe a three-layered model going from an outer layer to a core. The outer layer refers to observable behaviour, e.g. meetings, inspection reports, protective equipment, procedures. At a middle layer are espoused attitudes and perceptions, which are not directly observable but may be inferred from behaviour or by questioning. At the core level are

basic assumptions, e.g. about the nature of reality and truth, which are not easily assessed (ethnographic methods may be required).

A related three-level model of safety culture is proposed by Gordon et al. (2006). They regard safety culture as reflecting the actual level of commitment of norms set by safety management procedures. The safety culture model used specifies the relations between ‘what is believed’, ‘what is said’ and ‘what is done’. Discrepancies between these safety culture aspects are relevant for safety assessment and management.

In recognition of interactive contributions from psychological, situational and behavioural factors in accident causation, Cooper (2000) proposes a reciprocal safety culture model, which gives a multi-faceted view of safety culture (see Figure 13). In this approach, the internal psychological factors (i.e. attitudes and perceptions) are assessed via safety climate questionnaires, actual ongoing safety-related behaviour is assessed via checklists developed as a part of behavioural safety initiatives, and the situational features are assessed via safety management system audits/inspections.

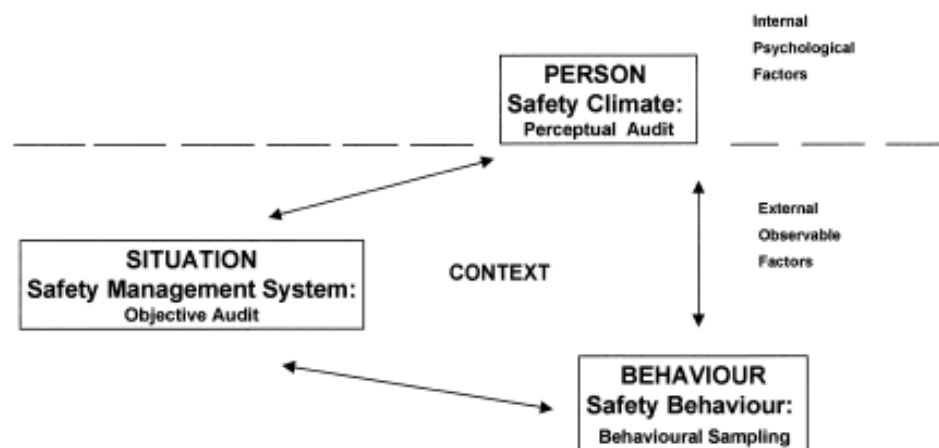


Figure 13: Reciprocal safety culture model (Cooper, 2000).

Ek et al. (2007) use a 9-dimensional model of safety culture, addressing 1) learning culture, 2) reporting culture, 3) just culture, 4) flexibility, 5) communication, 6) safety-related behaviours, 7) attitudes towards safety, 8) working situation, and 9) risk perception. They used this model in a survey-based analysis of safety culture in Swedish air traffic control and correlated the safety culture results with a 10-dimensional model of organizational climate. They found that the organizational climate dimensions *Support for ideas* (overall attitude towards new ideas) and *Conflicts* (presence of personal and emotional tensions) to be, respectively, positively and negatively related to many of the safety culture aspects.

Parker et al. (2006) provides a characterization of five levels of safety culture (from bad to good): pathological, reactive, calculative, proactive, generative. Part of the provided characteristics refers to concrete organizational aspects, e.g. auditing, incident/accident

reporting, contractor management, work-site job safety techniques. Other characteristics refer to ‘abstract’ organizational aspects, e.g. ‘how do safety meetings feel?’, balance between HSE and profitability, ‘who causes accidents in the eyes of the management?’.

Based on a review of safety culture literature, Montijn and De Jong (2006) identified six categories of safety culture aspects, which include 13 safety culture characteristics. Table 5 provides an overview of their results. In addition, they identified ranges of indicators for these characteristics, which may form the basis for questionnaires and interviews (Montijn and De Jong 2006).

Table 5: Categorized safety culture characteristics (Montijn and De Jong, 2006).

Category	Characteristics	Description
Commitment to safety	Attitudes towards safety	Reflects the extent to which the organization’s members and groups have a positive attitude towards the importance of safety and actually take personal responsibility for safe operations.
	Organizational commitment	Reflects the extent to which management identifies safety as a core value and guiding principle of the organization; demonstrates a positive attitude towards safety; and invests resources and effort in the implementation of safety improving activities.
	Employee empowerment	Reflects the extent to which (front-line) employees are empowered in keeping a high level of safety, by increased motivation and means in making the right decisions concerning safety issues.
Behaviour with respect to safety	Working situation	Reflects the extent to which the working situation promotes safe operations. The working situation includes team work, job pressure, job satisfaction, trust in colleagues and equipment.
	Safety related behaviours	Reflects the extent to which actions related to safety are taken by the organization’s members and groups.
	Management involvement	Reflects the extent to which management is involved in keeping a high level of safety throughout the entire organization.
Adaptability	Learning	Reflects the extent to which the organization’s members and groups are actively willing to learn from past experience regarding safety issues.
	Flexibility	Reflects the extent to which the organization is able to reconfigure itself when facing safety issues.
Awareness	Risk perception	Reflects the extent to which the organization’s members and groups are aware of the risk they and their surroundings are subjected to.
	Wariness	Reflects the extent to which the organization’s members and groups are maintaining a high degree of vigilance with respect to safety issues.
Information	Communication	Reflects the extent to which work related (thus not exclusively safety related) information is communicated to the right people within the organization.
	Reporting	Reflects the extent to which the organization’s members and groups are willing to report safety concerns, and to which they will not fear reprisals when doing so.
Justness	Justness	Reflects the extent to which safe behaviour and reporting of safety issues are encouraged or even rewarded, and unsafe behaviour is discouraged.

Assessment of safety culture

Kennedy and Kirwan (1998) developed the Safety Culture Hazard and Operability (SCHAZOP) approach that aims to identify (1) areas where the safety management process is vulnerable to failures, (2) the potential consequences of the safety management failures, (3) the potential safety culture failure mechanisms associated with the safety management failure, and (4) the factors which influence the likelihood of safety management failures. This is done with a HAZOP kind of approach following guide and property words, and discussion in expert groups of the listed issues.

Surveys (questionnaires, interviews) are often used to study organizational culture and their effect on safety (e.g. Ek et al. 2007, Montijn and De Jong 2006, Gordon et al. 2006). They can be used to study organizational practices, as well as attitudes. A drawback of surveys is that they tend to provide relatively superficial descriptions of organizational culture, since many practices are too complex and dynamic to be effectively captured in survey questions (Hopkins, 2006). Ethnographic research, where a researchers study the organization from within, can provide a much richer account of organizational culture than surveys can (Hopkins, 2006). Hopkins advocates the use of major accident inquiries for studying organizational culture and its impact on safety.

Reiman and Oedewald (2007) criticize the vague use of the term safety culture in organizational and safety research. According to them the concept of safety culture has become a catch-all concept for psychological and human factors issues in complex sociotechnical systems. The critique expresses a concern that safety culture is not seen as a contextual phenomenon, but as some kind of a general ideal model, without adequate consideration of the work itself that is being carried out in the organization in question. Reiman and Oedewald (2007) argue that

1. the current models of safety management are largely based on either a rational or a non-contextual image of an organization,
2. complex sociotechnical systems are socially constructed and dynamic cultures,
3. in order to be able to assess complex sociotechnical systems an understanding of the organizational core task is required, and
4. effectiveness and safety depend on the cultural conceptions of the organizational core task.

In relation to argument 3, Reiman and Oedewald (2007) argue that in order to make the theories of accidents and safety more contextual, understanding of the normal daily work, its objectives and its socially constructed nature are needed. Therefore, they introduce the concept of organizational core task (OCT), which refers to the shared objective or purpose of the organizational activity and describes the objective of the work, the characteristics of the physical object of the work and external influences (e.g. regulation, competition, suppliers). They define organizational culture as the process of formation and reformation of the conceptions concerning the organizational core task and the means to fulfil it.

For the cultural assessment in argument 4, Reiman and Oedewald (2007) distinguish three phases:

- a. Characterizing the culture of the organization. The aim is to exemplify the personnel's multiple ways of making sense of and interacting in the organizational context and to inspect what type of conceptions are shared among the personnel, and to what extent.
- b. Modelling the OCT in order to get appropriate criteria for the assessment of the organizational culture. The aim of the core task modelling is to extract demands of the work that apply to all the personnel. When analyzing the OCT, discrepancies in the conceptions of the organizational core task are considered to reflect the different aspects of the demands and the different angle from which the personnel perceive their organization and its core task.
- c. Explaining the effect of the culture on organizational effectiveness by qualitative assessment based on the OCT model and the extracted cultural features. Results include understanding reasons for different conceptions, identifying strengths and weaknesses of current practices, opening the dialogue on cultural aspects of work and effectiveness

Methods include interviews, surveys and working-groups, which are application dependent and provide qualitative results.

2.7 Human performance and human error

In many complex processes and organizations, and certainly in air traffic, human operators play key roles in providing the flexibility needed to deal effectively with the multiple, interrelated and often inherently uncertain processes. Given this key role it is no surprise that the human contribution to accident causation is usually considered to be significant. There exists a large volume of research on human factors and its relation to safe operations. This section provides an overview of some main lines in this research.

Human performance in an operational context

A well-known conceptual model for the performance of a human operator in the context of an operation is the SHELL model¹, which is depicted in Figure 14. It stresses the performance of a human ('Liveware') in the context of relations with other humans ('Liveware'), technical systems ('Hardware'), procedures ('Software') and the 'Environment' in which humans, technical systems and procedures must operate. The ragged sides of the blocks in the SHELL model in Figure 14 illustrate that the interfaces between the components may not precisely

¹ The SHELL (Software-Hardware-Environment-Liveware) concept was first developed by Edwards (1972) and basically describes the relations between human, machine and procedures in an environment. A three-dimensional SHELL model, which emphasizes the presence of several units of each type of SHELL resource and their interaction, is described in (Edwards 1988). A modified version of the original SHELL concept (Hawkins, 1987) uses a diagram (see Figure 14) that focuses on the human within its interaction with other humans, machines, procedures and environment. This SHELL diagram is currently used predominantly (e.g. ICAO 2002).

match and may induce tensions that may compromise human performance. In particular, the characteristics of each of the interfaces (Liveware-Hardware, Liveware-Software, Liveware-Liveware and Liveware-Environment) may have effect on (the variability of) the human performance.

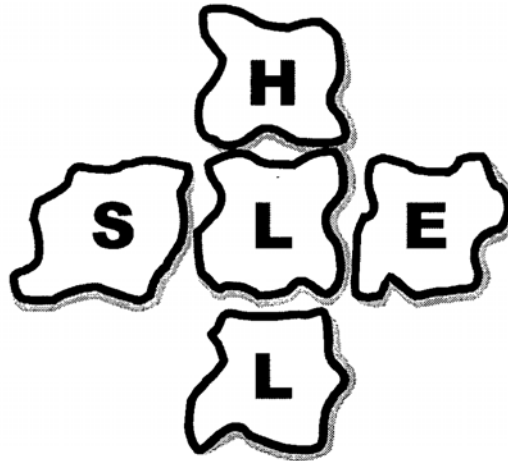


Figure 14: SHELL model, reflecting the working context of a human operator (Liveware): Software, Hardware, Environment and Liveware (ICAO 2002; based on a model by Hawkins 1987).

Generic human factors that determine human capability to deal with incongruent interfaces are

- physical factors, i.e. human physical capabilities to perform tasks, e.g. strength, vision, hearing, and tolerance to heat, light, noise, vibration;
- physiological factors, i.e. factors affecting physiological processes and thereby restricting physical performance, e.g. oxygen availability, general health, alcohol use, fatigue;
- psychological factors, i.e. factors affecting the psychological preparedness to meet the task demands, e.g. motivation, confidence, risk taking behaviour attitude, stress;
- psycho-social factors, i.e. factors from the social environment of a human, e.g. labour-management disputes, family problems, financial problems.

The SHELL model describes factors influencing the performance of a single human operator (including human-human interactions). In a multi-agent organization there are multiple human operators, each working in its own SHELL. The SHELL components of different human operators may be unique, partly overlapping or (almost) identical. Furthermore, the SHELL components and its interfaces are dynamic and their dynamics may be interdependent.

Human reliability assessment

Historically, most safety research and applications have focused on sequential and, more recently, epidemiological accident models (Holnagel, 2004; see also Section 2.8). In describing accident occurrence, they both use cause-effect propagation for technical systems as well as



human operators. For technical systems this usually comes down to evaluation of the effects of failures and system redundancy, for human operators it typically implies the evaluation of human error and the potential of human operators to resolve safety-critical situations. In this context, human performance is evaluated by human reliability assessment (HRA). Its name clearly elucidates the focus on human failure and reflects the similar way of representing performance of humans and technical systems via reliability in safety assessment. The main steps in HRA are (Kirwan, 1994):

1. *Problem definition.* What human involvements are to be assessed, e.g. emergency situations, maintenance procedures, etc.?
2. *Task analysis.* What is an operator required to do, in terms of actions and/or cognitive processes, in the situations under analysis?
3. *Human error identification.* What can go wrong in the human performance? Identification of failures related to human performance, such as errors of omission (failing to carry out a required act), errors of commission (failing to carry out an act adequately, e.g., lack of precision, wrong timing, wrong sequence) and extraneous act (unrequired act performed instead of, or in addition to, required act), as well as identification of error-recovery opportunities. Such external manifestations of an error are known as external error modes (EEM). In addition internal manifestations of errors may be identified, e.g. memory failure, pattern recognition failure, vigilance failure, confirmation bias, cognitive overload. Such internal manifestations of an error are known as psychological error mechanism (PEM). The PEM may be described at various levels, e.g. the terms error detail, error mechanism, information processing level are used in HERA-JANUS (Isaac et al., 2003).
4. *Representation.* How can the acquired information be represented for quantitative evaluation of human-error impact in the context of other potential contributions to system risk, such as hardware and software failures, and environmental events? HRA often applies fault and event trees for this.
5. *Human error quantification.* Quantification of the likelihood of error and evaluation of the overall effect of human error on system safety and reliability. Human reliability quantification techniques all involve evaluation of human error probability (HEP), being the ratio of the number of times an error has occurred and the number of opportunities for that error to occur. The effect of contextual conditions on the HEP is usually evaluated via Performance Shaping Factors (PSFs) or Error Producing Conditions (EPCs) (Gertman and Blackman, 1994; Kirwan et al., 2004). They indicate how contextual conditions, such as level of stress, level of training and quality of man-machine interface, influence the likelihood of error occurrence.
6. *Impact assessment.* Evaluation of the overall risk, the acceptability of the risk and determination of most important risk contributors.



7. *Error reduction analysis*. Proposal and analysis of error reduction measures to reduce the error likelihood and/or impact.
8. *Documentation and quality assurance*. Documentation of results and quality assurance to ensure that required error reduction measures are effectively implemented and assumptions made remain valid.

A range of techniques has been developed to support the steps in HRA. Appendix A provides descriptions of some prominent ones. Well known techniques for human error identification in the context of air traffic are HFACS (Wiegmann and Shappel 2001; Appendix A.1) and Tracer/HERA (Shorrock and Kirwan 2002, Isaac et al. 2003; Appendix A.2). Prominent examples of techniques for human error quantification include Technique for Human Error Rate Prediction (THERP; Swain and Guttman 1983) and Human Error Assessment and Reduction Technique (HEART; Williams 1988; Appendix A.3). General risk assessment techniques, which are also used extensively for human error evaluation, are expert judgement (Appendix A.4) and the use of accident/incident databases (Appendix A.5).

Short-coming of HRA methods have been recognized for quite some time now (Dougherty 1990, Swain 1990) and address issues such as

- the lack of sufficient human performance data for quantitative prediction (of human error),
- the validity of expert judgements for rare events,
- inconsistency between simulator data and real life,
- lack of psychological realism in HRA methods,
- the validity of performance shaping factors for accounting for contextual conditions.

In an attempt to address these kinds of criticisms, so-called 'Second Generation HRA' methods have been developed, such as A Technique for Human Error Analysis (ATHEANA; Cooper et al. 1996) and Cognitive Reliability and Error Analysis Method (CREAM; Hollnagel 1998); see also (Sträter 2005) for a further discussion of First and Second Generation methods. These methods propose dedicated ways to incorporate the effects of contextual conditions on human performance and the likelihood of something being done incorrectly. Notwithstanding the dedicated ways to handle the effect of contextual conditions in a Second Generation HRA method like CREAM, in the end it gives probabilities for cognitive function failures, which are integrated in a tree-based probabilistic safety assessment approach. Therefore, this kind of method confers to the line of reasoning in sequential and epidemiological accident models (see Section 2.8).

Human performance in systemic accident modelling

Systemic accident models focus on the variability in the performance of human operators. In connection with the variability of related agents (humans / systems), safety critical events may emerge from this variability (Hollnagel 2004; see also Section 2.8). Systemic accident modelling

represents a new way of viewing accident causation and supporting techniques are being developed. Specific techniques for systemic accident modelling are discussed in Section 2.8.

Regarding the inclusion of human performance, the principle of systemic accident modelling favours the use of broad-scope models for human performance. Examples of human performance aspects that may be included in systemic accident modelling include the integrated approach for human factors developed by Eurocontrol named Human Factors Case (Shorrock et al. 2004, Mellett and Nendick 2007; see also Appendix A.6), situation awareness (Endsley 1995), multi-agent situation awareness (Stroeve et al. 2003) and the contextual control mode model (Hollnagel 1998). Also specific HRA techniques may be valuable in systemic accident modelling. For instance, in some cases it may still be useful to distinguish human actions with undesired consequences and evaluate probabilities for them. Quantification in this approach may be done for a broad range of human performance aspects, e.g. probability density functions of task duration, inter-monitoring times, velocity estimation accuracy or event occurrences.

Team processes

Grote et al. (2004) describe characteristics of organizations, teams and tasks that influence team processes. Their model distinguishes three levels, each including multiple characteristics:

1. Organizational aspects, encompassing
 - a. Regulatory influences, e.g. prescriptive regulation or goal-oriented legislation;
 - b. Organizational culture, which has relations with national culture, professional culture and safety culture;
2. Task characteristics, encompassing
 - a. Dynamics and automation of the task within the system;
 - b. Dependency on verbal communication;
 - c. Uncertainty of disturbances;
 - d. Context set by economic, organizational and regulatory constraints;
3. Individual characteristics, encompassing
 - a. Task familiarity: skill / rule / knowledge based;
 - b. Cultural communication skills;
 - c. Individual goals and attitudes;
 - d. Team composition: distribution of knowledge and experience.

2.8 Accident models incorporating organizational aspects

2.8.1 Problem

In complex and distributed organizations the level of safety is the result of interactions between many entities of various types at multiple locations. In organizational risk assessment the roles of these various entities may be elucidated by using the sharp end – blunt end concept

(Hollnagel 2004). The sharp end refers to the people who are working at the time and in the place where the accident may take place, e.g. pilots, physicians, controllers, power plant operators. The blunt end refers to the people affecting safety through their effect on the constraints and resources acting on the practitioners at the sharp end, e.g. regulators, managers.

For the analysis of organizational accidents a well known concept popularized by Reason (1997) is that latent conditions produced by people working at the blunt end may erode safety barriers of people working at the sharp end (see Figure 15). A related way to view this is by the conceptual Swiss Cheese model of Reason (1997), which depicts that multiple contributors (the holes in the cheese slices) must be aligned to circumvent the safety barriers (the cheese slices) such that an accident occurs.

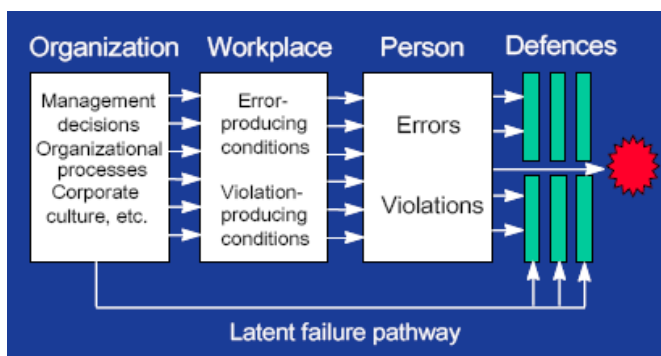


Figure 15: Latent conditions affecting safety defences (Reason et al. 2006)

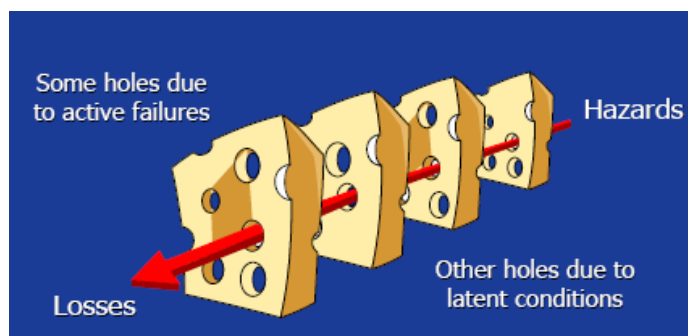


Figure 16: Swiss cheese model of accident causation (Reason et al. 2006).

Messages similar to those of Reason (1997) have been put forward by Turner (1978) and Pidgeon and O’Leary (2000) in their man-made disasters theory. The man-made disasters theory states that, despite the best intentions of all involved, the objective of safely operating technological systems could be subverted by some very familiar and ‘normal’ processes of organizational life. A disaster is defined in the man-made disasters model not by its physical impacts at all, but in sociological terms, as a significant disruption or collapse of the existing cultural beliefs and norms about hazards, and for dealing with them and their impacts. Man-made disasters theory highlights how system vulnerability often arises from unintended and

complex interactions between contributory preconditions, each of which would be unlikely, singly, to defeat the established safety systems. A further key part of the organizational aetiology of disaster incubation is the way in which the order-producing tendencies of social systems contribute to the generation of extreme hazard from relatively safe situations, through the structured amplification of the consequences of earlier errors.

Also Perrow (1984) provided early descriptions of accidents as the consequence of complex interactions in sociotechnical systems and considers such accidents to be normal: “If interactive complexity and tight coupling – system characteristics – inevitably will produce an accident, I believe that we are justified in calling it a *normal accident*, or a *system accident*. The odd term *normal accident* is meant to signal that, given the system characteristics, multiple and unexpected interactions of failures are inevitable. This is an expression of an integral characteristic of the system, not a statement of frequency.”

Le Coze (2005) argues that organisations are complex systems, which have numerous interactions, non-linear cause-effect relationships, unclear boundaries, adaptive features and complex interactions between individuals, depending on cognitive, cultural, power, rules, social construction, etc. Because of this complexity, organisations cannot be studied properly by traditional technical risk assessment approaches. As such, Le Coze (2005) stipulates that for safety auditing of organisations we need to (1) acknowledge the organisational complexity, (2) study the multi-dimensionality of organisational aspects at various levels with multiple disciplines, and (3) learn from detailed accident analyses.

2.8.2 Types of accident models

In arguing about the occurrence of accidents, we use a frame of reference about how accidents may come about. Hollnagel (2004) distinguishes three types of such accident models: sequential accident models, epidemiological accident models and systemic accident models; they are described next.

Sequential accident models

Sequential accident models describe an accident as the result of a sequence of events that occur in a specific order (Hollnagel 2004). These models assume that there are well-defined cause-effect links that propagate the effects of events leading to an accident. Sequential accident models are used extensively in risk assessment, examples include the domino theory, event trees, fault trees and networks models. Figure 17 shows a combination of a fault and event tree such as used in the Eurocontrol Safety Assessment Methodology (Eurocontrol 2004a).

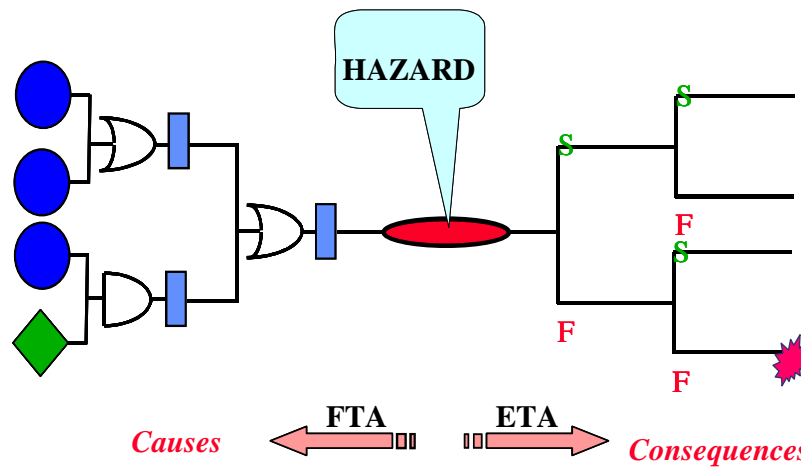


Figure 17: Example of a sequential accident model: fault and event trees (Eurocontrol 2004a).

Epidemiological accident models

Epidemiological accident models describe an accident in analogy with the spreading of a disease, i.e. as the outcome of a combination of factors, such as performance deviations, environmental conditions, barriers and latent conditions (Hollnagel 2004). Like sequential accident models, epidemiological accident models rely on cause-effect propagation in accidents. Examples of epidemiological models are the “Swiss cheese” model of Reason (see Figure 16) and Bayesian belief networks (see Figure 18). Epidemiological models provide a broader basis to represent the complexity of accidents than sequential models by accounting for more complex interactions between relevant factors. For instance, the example of Figure 18 accounts for conditional probability distributions of factors contributing to various types of accidents.

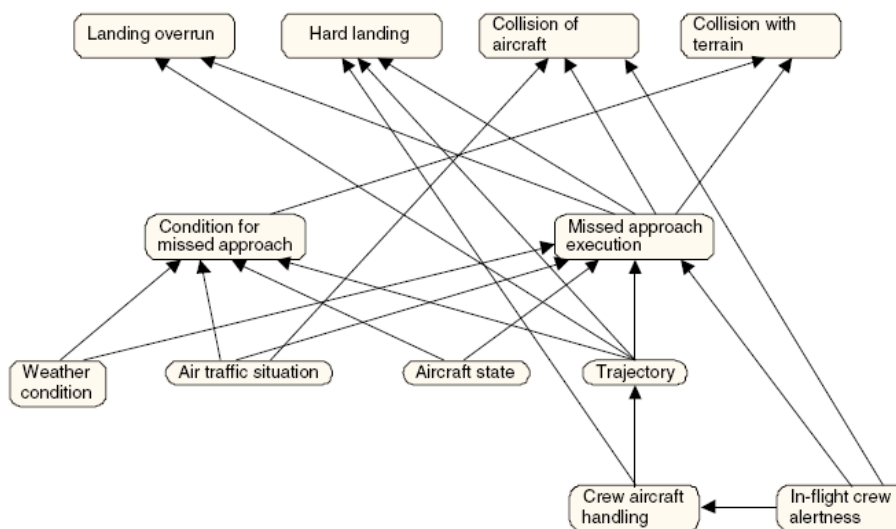


Figure 18: Example of an epidemiological accident model: Bayesian belief network (Ale et al. 2006).

Epidemiological accident models for organizational processes include Tripod (Hudson et al. 1994, Groeneweg 1998; see Appendix A.7) and the Markov Latent Effects approach (Cooper 2001). The Markov Latent Effects approach aims at the quantification of safety effects of organizational and operational factors that can be measured through “inspection” or surveillance. It describes relations between organizational aspects and integrates them to the level of safety. For these relationships it uses imprecise metrics in possibilistic or fuzzy numbers to quantify the quality of organizational aspects. It uses deterministic weighting parameters to represent the impact of organizational aspects in an overall architecture of the organization and associated risk factors.

Both sequential and epidemiological accident models rely on cause-effect propagation in accidents and give a fixed representation of relations between failures, errors and contextual conditions. Predominantly, they represent relations between probabilities of event occurrences that are cause-effect implied. The boundary between these two types of accident models is not very sharp and in practice accident models may have characteristics of both. For instance, the Intergrated Risk Picture for ATM in Europe (Spouge and Perrin 2006) combines a fault tree (which is a sequential accident model) with an influence model that addresses causal factors influencing bottom event in the fault tree, which has the characteristics of an epidemiological accident model.

Recently, Reason et al. (2006) have stated that there is a need for models that can account for accidents as due to conjunctions of variability (i.e. systemic accident models (Hollnagel 2004)). They acknowledge that the Swiss Cheese Model is not a detailed accident model describing how the multitude of functions and entities in a complex socio-technical system interdependently interact. Nevertheless it has had a significant influence on the understanding of accidents and still is a valuable means of communication and a heuristic explanatory device (Reason et al. 2006).

Systemic accident model

Systemic accident models describe the performance of a system as a whole, rather than on the level of cause-effect mechanisms or epidemiological factors (Hollnagel 2004). The systemic accident model view considers accidents as emergent phenomena from the performance variability of a system, in particular due to the dynamic interaction between multiple agents (humans, technical systems), which form a joint cognitive system. A joint cognitive system is a combination of two or more systems, where at least one is a cognitive system (e.g. a human) (Hollnagel and Woods 2005). The performance of technical systems and human operators is to some extent always variable. This variability is due to noise influences on the performance and due to the interactions between entities in a complex organisation. In fact, Hollnagel (2004) argues that human performance must be variable to handle efficiently the complex interactions in a socio-technical environment. Hollnagel also argues that the combined and coupled

performance variability in an organisation may lead to functional resonance, i.e. enlarged deviations in performance from normal practice. In this framework, considerable functional resonance is seen as the source of accident causation. Pariès (2006) points out that to understand the properties of a complex system, we lay relationships between micro and macro levels, such that macro level properties emerge from assembling micro level properties. These views indicate that for risk assessment of complex organisations, of which air traffic is a prime example, we need analysis approaches that account for the variability in their multi-agent performance and the emergence of safety occurrences from this variability.

Systemic accident models have their origins in cybernetic control theory and chaos theory. Recent developments in systemic accident modelling include ‘Organizations in context’ approach, Functional Resonance Accident Model (FRAM), Systems-Theoretic Accident Model and Processes (STAMP), and Traffic Organization and Perturbation AnalyZer (TOPAZ).

- The ‘Organizations in context’ approach of Dyhrberg and Jensen (2004) builds on the risk management model of Rasmussen and Svedung (2000). It encompasses the use of a framework of contextual approaches for analysis of contextual influences on occupational accidents. These approaches are characterised by viewing an organization as a complex entity where behaviour and functions depend on actors and agencies outside the organization. The approaches include decision-making theory, regulatory theory (on the effect of regulation), relational theory (on types of relations between agents) and institutional theory (on the impact of socially constructed rules and norms, e.g. culture).
- FRAM uses a functional representation of an operation and functional diagrams, which are an extension of the Functional Analysis and Design Technique (SADT) (Hollnagel 2004). FRAM describes performance variability based on a number of characteristics of each function (input, output, resources, time, control and preconditions) and the interactions with other functions. A qualitative analysis is used to evaluate safety-critical conditions where the interdependencies in a FRAM network may give rise to functional resonance.
- STAMP is a model based on system and control theory, and uses the key principle that accidents may occur as the result of a lack of control constraints imposed on the system design and on operations (Leveson 2004; Appendix A.8). STAMP applications have mostly focussed on interactions at higher organisational levels than the human-system level. STAMP supports quantitative evaluation of risk levels as function of organisational processes, but not at the level of accident probability.
- The TOPAZ methodology is based on stochastic system and control theory, and employs the basic principle that accidents may result from stochastic and dynamic interactions between agents in a traffic scenario (Blom et al. 2001a; Appendix A.9). It uses integration with a qualitative safety risk assessment cycle, mathematical modelling, Monte Carlo simulation and uncertainty evaluations to analyse the safety of air traffic operations up to the level of accident probability. Multi-agent situation awareness is a key concept in TOPAZ.

3 Evaluation of potential application cases

In this project the types of formal organizational modelling techniques presented in Section 2 will be evaluated for their use in safety assessment of organizational processes in air traffic. To this end an application case is selected for which the formal organizational modelling will be evaluated. The current section describes the selection process of the application case.

In coordination with Eurocontrol the following potential applications cases have been identified.

1. *Taxiing operations and runway incursion risk.* It concerns the organization of taxiing on the aerodrome and the risk of runway incursion.
2. *Safety management at Eurocontrol.* The safety management at Eurocontrol is organised in a ‘virtual safety team’ with contributions from its locations in Brussels, Brétigny, Luxembourg, Maastricht. Within this case inter-organisational cooperation processes may be modelled and analysed with the aim of identifying bottlenecks and inefficiencies and proposing possible solutions.
3. *Safety culture and incident reporting.* Organisation of incident reporting in an ANSP and its effect on safety culture and safety management.
4. *Free flight.* The organization of delegation of separation responsibilities from ANSP’s to aircraft crews supported by appropriate on-board systems.

Table 6: Evaluation criteria for the selection of an application case.

ID	Name	Explanation
C1	Relevance	The application should be relevant for air traffic safety.
C2	Complexity	The application should be sufficiently complex to illustrate the value of formal analysis.
C3	Knowledge	There should be sufficient domain knowledge readily available or easily obtainable on the organizational structure of the application.
C4	Current models	The existence of safety models of the operation considered in the organizational application would be an advantage.
C5	Hazards	Knowledge of hazards in the application would be an advantage.
C6	Pragmatic	The potential feasibility and usability must be shown within the first year.

For evaluation of these potential application cases the criteria presented in Table 6 have been identified. The results of the evaluation are shown in Table 7. Based on this evaluation, option 1: ‘Taxiing operations and runway incursion risk’ is considered most suitable for the follow-up work. Incident reporting and its effects on safety culture/management (option 3) may be

considered within the context of the organization of taxiing operations and runway incursion risk.

Table 7: Evaluation of potential application cases. The score for each criterion ranges from - - (not suitable) to ++ (very suitable).

Criterion		Evaluation	
ID	Name	Score	Explanation
<i>Option 1: Taxiing operations and runway incursion risk</i>			
C1	Relevance	++	Runway incursion is considered to be a prime safety issue in air traffic. Several safety programmes have been initiated to restrict the risk of runway incursion, e.g. (Eurocontrol 2004b).
C2	Complexity	++	Aerodrome operations are considerably complex, involving many agents from various organizations.
C3	Knowledge	+	NLR has considerable domain knowledge on taxiing operations. This knowledge would have to be structured for this organizational context. Additional information may be needed on processes at higher levels of the organization.
C4	Current models	++	NLR has various accident models for runway incursion risk, which have been applied for risk assessment of actual operations.
C5	Hazards	+	A hazard database for hazards at the operational level is available at NLR. Hazards at higher organizational levels may need to be identified.
C6	Pragmatic	+	It may be necessary to restrict the complexity in the application case to a number of relevant organizations and agents.
<i>Option 2: Safety management at Eurocontrol</i>			
C1	Relevance	+	The organization of safety management at Eurocontrol is relevant for air traffic safety. Being at a distance from the actual performance of air traffic operations, it has mostly an indirect effect on their safety.
C2	Complexity	0	Eurocontrol's safety management involves 'virtual safety teams' at various locations.
C3	Knowledge	-	Information on the organization may be obtained from organigrams, job descriptions and interviews. It may be difficult to obtain appropriate information given potential

			variety of interests.
C4	Current models	--	There are no current models.
C5	Hazards	--	Hazards have not yet been identified.
C6	Pragmatic	-	The required inter-organizational analysis and expected level of politics add to the level of complexity.
<i>Option 3: Incident reporting and safety culture/management</i>			
C1	Relevance	+	Incident reporting is relevant for air traffic safety. The relation between safety culture/management and the achieved level of safety is somewhat oblique when considered in general terms. Such relations can best be clarified by considering the organization in a specific operational context.
C2	Complexity	?	The complexity depends on the operational and organizational contexts studied.
C3	Knowledge	?	The availability of organizational knowledge depends on the operational and organizational contexts studied.
C4	Current models	-	Research on formal models is in an early stage.
C5	Hazards	-	Hazards in relation to incident reporting and safety culture/management have not been systematically identified. Specific hazards depend on the operational and organizational contexts studied.
C6	Pragmatic	?	Whether this option may be pragmatic would depend on the operational and organizational contexts studied.
<i>Option 4: Free flight</i>			
C1	Relevance	+	Free flight being a concept for future operations, it is relevant for the development of air traffic and the safety of future operations.
C2	Complexity	++	Free flight operations are very complex.
C3	Knowledge	0	At the operational level, concepts for free flight have been developed, which display various views on the distribution of roles and responsibilities. Being concepts of future operations, the imbedding in organizational processes is part of the development.
C4	Current models	+	NLR has accident models for free flight operation. Since free flight is not yet operational, existing safety models are somewhat limited.
C5	Hazards	+	A hazard database for hazards at the operational level is



			available at NLR. Hazards at higher organizational levels may need to be identified.
C6	Pragmatic	--	Given the lack of knowledge on the organization of free flight this option is considered not pragmatic.

4 Concluding remarks

The objective of this research project is to enhance safety analysis of organizational processes in air traffic by development of formal approaches for modelling, simulation and analysis of organizational relationships and processes. As a basis for this research the current report has presented a wide overview of related methods, techniques and research identified in the literature.

Referring to the organizational safety pyramid introduced in Figure 1 on page 10, the identification of safety issues and evaluation of risk levels for organizational processes can be described in a systemic accident model context at three levels: (1) organizational structure, (2) performance variability and (3) organizational safety.

The first level describes the organizational structure, which encompasses human and technical organizational entities and various types of relations (e.g. power, communication) between these entities that occur in the executions of organizational work processes. The organizational structure may be specified at different aggregation levels ranging from general regulations for the whole organization to specific prescriptions for particular roles and their interaction with other roles. A simple connotation for this layer on organizational structure is a slight variation on the basic definition of organizational culture “the way we do things around here”, namely “the way we should do thing around here”. It thus focuses on work processes and relations between humans and technical systems as they are envisioned within the organization. Analysis of processes and relations both at the same aggregation level and across different levels may show (potentially safety-critical) inconsistencies.

Multi-agent models cross the boundary between the first and second level in the organizational safety pyramid. On the one hand, their interactions are regulated by the organizational structure. On the other hand, the dynamic and stochastic aspects of interacting agents contribute to the performance variability in an organization.

The second level describes key aspects of sources for performance variability in an organization. As prime source of performance variability it addresses human performance. Human performance is flexible and adaptable, work may be done faster/slower, tasks may be done more or less precisely, external cues may be detected or not, tasks may be omitted or done in various orders, etc. Within a systemic accident model this variability should be described to an extent appropriate for the role of the human operator within the blunt side or sharp side in the organization. The organizational or safety culture can be seen at the second level as an important moderator of the extent of human performance variability and the quality of interactions between organizational entities. For instance, the organizational culture may be such that control checks are frequently omitted, high work loads are considered normal or problems are not communicated within the organization. In this view organizational culture (including norms, values, etc.) is made specific by indicating its effect on (the variability in) the



performance of humans. In total, this thus leads to a model of “the way we do things around here”.

The third level describes how the variability in performance of the humans, the variability in the performance of technical systems, and the variability in the performance of interacting humans, interacting humans and technical systems, and interacting technical systems may lead to inconsistencies, problems, incidents and accidents. It is thus at this level that accident causation may actually be described as resulting from the dynamic and stochastic interactions between the agents in the organization.

Key within the presented organizational safety pyramid is the local perspectives and behaviour of separate actors in organizational structures. Multi-agent models provide a suitable framework to represent these actors and their interactions. Accidents are considered as emergent phenomena from the variability of the agents’ performance in the organizational context. In multi-agent systems conflicts can be discerned at various levels: (1) at a physical level, e.g. multiple aircraft striving to use a runway as soon as possible, (2) at a knowledge level, e.g. pilots and controllers using different information, or (3) at an organizational level, e.g. organizational norms and prescriptions imposed on the agents may be in conflict with their mental attitudes (goals, wishes, desires etc.). For the safety of complex air traffic organizations, understanding of the emerging of conflicts between its constituent agents is of key interest.

This organizational safety pyramid fits well within the systemic accident model view, which considers accidents as emergent phenomena from the variability of a system (or organization) as a whole (Hollnagel 2004). Using the three level view in this pyramid, in the remainder of this research feasible ways will be identified as to how the models at the first level can be used to further support safety research at the second and third levels.

References

- Abbink H, Van Dijk R, Dobos T, Hoogendoorn M, Jonker CM, Konur S, Van Maanen PP, Popova V, Sharpanskykh A, Van Tooren P, Treur J, Valk J, Xu L, Yolum P (2004). Automated Support for Adaptive Incident Management, In: Van de Walle B, Carle B (eds.), *Proceedings of the First International Workshop on Information Systems for Crisis Response and Management*, ISCRAM'04, pp. 69-74
- Ale BJM, Bellamy LJ, Cooke RM, Goossens LHJ, Hale AR, Roelen ALC, Smith E (2006). Towards a causal model for air transport safety - an ongoing research project. *Safety Science*, Volume 44, Issue 8, Pages 657-673
- Bacharach, SB., Lawler, EJ (1980) *Power and politics in organizations*, Jossey-Bass, San Francisco
- Biddle B (1979) *Role Theory: Concepts and Research*, Krieger Publishing Co.
- Blau PM, Schoenherr RA (1971). *The structure of organizations*, Basic Books Inc., New York London
- Blom HAP, Bakker GJ, Blanker PJG, Daams J, Everdij MHC, Klompstra MB (2001a). Accident risk assessment for advanced air traffic management. In: Donohue GL and Zellweger AG (eds.), *Air Transport Systems Engineering*, AIAA, pp. 463-480
- Blom HAP, Daams J, Nijhuis HB (2001b). Human cognition modelling in air traffic management safety assessment. In: Donohue GL and Zellweger AG (eds.), *Air Transport Systems Engineering*, AIAA, pp. 481-511
- Blom HAP, Bakker GJ, Everdij MHC, Van der Park MNJ (2003a). Collision risk modelling of air traffic. *Proceedings European Control Conference 2003 (ECC03)*, Cambridge, UK
- Blom HAP, Bakker GJ, Everdij MHC, Van der Park MNJ (2003b). *Stochastic analysis background of accident risk assessment for air traffic management*. Hybridge EC project, D2.2
- Blom HAP, Stroeve SH, Everdij MHC, Van der Park MNJ (2003c). Human cognition performance model to evaluate safe spacing in air traffic, *Human Factors and Aerospace Safety*, Vol. 3, pp. 59-82
- Blom HAP, Klompstra MB, Bakker GJ (2003d). Accident risk assessment of simultaneous converging instrument approaches. *Air Traffic Control Quarterly* 11: 123-155
- Blom HAP, Stroeve SH (2004). Multi-agent situation awareness error evolution in air traffic. *Proceedings 6th Probabilistic Safety Assessment and Management Conference*, Berlin, Germany
- Blom HAP, Stroeve SH, De Jong HH (2006a). Safety risk assessment by Monte Carlo simulation of complex safety critical operations. In Redmill F, Anderson T (eds.), *Developments in risk-based approaches to safety*, Springer-Verlag, London

- Blom HAP, Bakker GJ, Klein Obbink B, Klompstra MB (2006b). Free flight safety risk modeling and simulation. *Proc. Int. Conf. on Research in Air Transport (ICRAT)*, Belgrade, Servia
- Bosse T, Jonker CM, Meij L, Treur J (2005). LEADSTO: a Language and Environment for Analysis of Dynamics by SimulaTiOn. In *Proceeding of the Third German Conference on Multi-Agent System Technologies, MATES'05*, Lecture Notes in Artificial Intelligence, vol. 3550, Springer Verlag, 165-178
- BPML. *Business Process Modeling Language (BPML)*. <http://www.bpml.org>
- Brazier, F.M.T., Dunin-Keplicz, B., Jennings, N., and Treur, J. (1997) DESIRE: Modelling Multi-Agent Systems in a Compositional Formal Framework. *International Journal of Cooperative Information Systems*, 6: 67-94
- Bratman, M. E. (1999). *Intention, Plans, and Practical Reason*. CSLI Publications
- Bresciani P, Giorgini P, Giunchiglia F, Mylopoulos J, Perini A (2004). Tropos: An Agent-Oriented Software Development Methodology, *Journal of Autonomous Agent and Multi-Agent Systems*, vol. 8(3): 203-236
- CAA (Civil Aviation Authorities) (1993). *Hazard analysis of an en-route sector*, Volumes 1 and 2. RMC Report R93-81(S)
- Cammarata, S., McArthur, D., and Steeb, R. (1983). Strategies of cooperation in distributed problem solving. In *Proceedings of the International Joint Conference of Artificial Intelligence '83*, 767-770.
- Campion MA, Medsker GJ, Higgs AC (1993). Relationship between Work Group Characteristics and Effectiveness: Implications for Designing Effective Work Groups. *Personnel Psychology*, 46: 823-850
- Carley KM (1996). A comparison of artificial and human organizations. *Journal of Economic Behavior & Organization*, 31(2): 175-191
- Carley, KM, Kjaer-Hansen, J, Newell, A. and Prietula, M. (1991) Plural-SOAR: capabilities and coordination of multiple agents, in: Masuch, M. and, Warglien M. (eds.), *Artificial intelligence in organization and management theory*, Elsevier Science.
- CD 14258 "Industrial automation systems - Rules and guidelines for enterprise models" ISO TC184/SC5/WG1, 1996
- CEN/TC310 "CIM Systems Architecture - Enterprise model execution and integration services - Evaluation report" CEN Report CR: 1831 :1995
- CEN/TC310 "CIM Systems Architecture - Enterprise model execution and integration services - Statement of Requirements" CEN Report CR: 1832 :1995
- Checkland P, Holwell S (1998) *Information, Systems and Information Systems - making sense of the field*, John Wiley & Sons
- Child J (1973) Organization: A Choice for Man., In Child J (ed), *Man and Organization*, Halsted Press, London, 234-570

- Choudhry RM, Fang D, Mohamed S (2006). The nature of safety culture: A survey of the state-of-the-art. *Safety Science* doi10.1016/j.ssci.2006.09.003 (in press)
- CIMOSA (1993). *CIMOSA – Open System Architecture for CIM*; ESPRIT Consortium AMICE, Springer-Verlag, Berlin
- Cohen, GP (1992) *The Virtual Design Team: An Information Processing Model of Coordination in Project Design Teams*. PhD Thesis, Stanford University, CA
- Cooke RM, Goossens LHJ (2000). *Procedures guide for structured expert judgement*. EUR 18820 EN
- Cooper MD (2000). Towards a model of safety culture. *Safety Science* 36: 111-136
- Cooper JA (2001). *The Markov latent effects approach to safety assessment and decision-making*. Sandia National Laboratories, report SAND2001-2229
- Cooper SE, Ramey-Smith AM, Wreathall J, Parry GW, Bley DC, Luckas WJ, Taylor JH, Barriere MT (1996). *A technique for human error analysis (ATHEANA) - Technical Basis and Method Description*. Nureg/CR-6350. US Nuclear Regulatory Commission, Washington D.C.
- Cox, S., Cox, T. (1996). *Safety, Systems and People*. Butterworth-Heinemann, Oxford
- Cummings TG, Worley CG (2005) *Organization development and change*. Thomson/South Western
- Davis, R., Smith, R. (1983) Negotiation as a metaphor for distributed problem solving. *Artificial Intelligence*, 20:63-109
- Decker, K. and Lesser, V. (1992) Generalizing the Partial Global Planning Algorithm. *International Journal on Intelligent Cooperative Information Systems*, Vol. 1, Number 2, pp. 319-346. June 1992
- De Jong HH, Stroeve SH, Blom HAP (2006). The roles of air traffic controllers and pilots in safety risk analyses. *ESREL Conference*, 18-22 September 2006, Estoril, Portugal
- Demazeau, Y. (1995) From Interactions to Collective Behaviour in Agent-Based Systems. In: *Proceedings of the 1st. European Conference on Cognitive Science*. Saint-Malo, France.
- d’Inverno, M., Kinny, D., Luck, M. (1997) A formal specification of dMARS. In: *Proceedings of the Fourth International Workshop on Agent Theories, Architectures, and Languages, ATAL’97*. Auch: AAIS - Technical Note 73, Australian Artificial Intelligence Institute, Melbourne
- Donaldson L. (2001) *The Contingency Theory of Organizations*. Sage, London
- Dougherty E.M. (1990) Human Reliability Analysis – Where shouldst thou turn? *Reliability Engineering and System Safety* 29:283-299
- Doumeingts G, Vallespir B, Chen D (1998). Decisional Modelling using the GRAI Grid, In : Bernus, P., Mertins, K. and Schmidt, G. (Eds): *Handbook on Architectures of Information Systems*, Springer-Verlag 313-338

- Durfee, EH, Lesser, VR and Corkill, DD (1990) Cooperation Through Communication in a Distributed Problem-Solving Network. *Cognition, Computing, and Cooperation*, S. Robertson, W. Zachary, and J. Black, ed., Ablex Publishing Company, pp. 159-186
- Dyhrberg MB, Jensen PL (2004). Organizations in context: proposal for a new theoretical approach in prescriptive accident research. *Safety Science* 42:961-977
- Edwards E (1972). Man and machine: systems for safety. *Proceedings of the BALPA Technical Symposium*, London
- Edwards E (1988). Introductory Overview. In Wiener EL, Nagel DC (eds.), *Human factors in aviation*. Academic Press, San Diego (CA), USA
- Ek A, Akselsson R, Arvidsson M, Johansson CR (2007). Safety culture in Swedish air traffic control. *Safety Science* 45(7):791-811
- Endsley MR (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors*, 37(1): 32-64
- ENV 40 003 Computer Integrated Manufacturing - Systems Architecture - Framework for Enterprise Modelling CEN/CENELEC, 1990
- ENV 12 204 Advanced Manufacturing Technology - Systems Architecture - Constructs for Enterprise Modelling CEN TC 310/WG1, 1995
- Epstein, JM and Axtell, R. (1997) *Growing Artificial Societies*. MIT Press, Cambridge, MA.
- Everdij MHC, Blom HAP, Stroeve SH (2006a). Structured assessment of bias and uncertainty in Monte Carlo simulated accident risk. *Proceedings of the 8th International Conference on Probabilistic Safety Assessment and Management*, May 14-18 2006, New Orleans, USA
- Everdij MHC, Klompstra MB, Blom HAP, Klein Obbink B (2006b). Compositional specification of a multi-agent system by stochastically and dynamically coloured Petri nets. In: Blom HAP, Lygeros J (eds.), *Stochastic Hybrid Systems*, LNCIS 337, Springer-Verlag, pp. 325-350
- Eurocontrol (2004a). *Air navigation system safety assessment methodology*. SAF.ET1.ST03.1000-MAN-01, edition 2.0
- Eurocontrol (2004b). *European action plan for the prevention of runway incursions*, release 1.1
- Ferber J, Gutknecht O (1998). A meta-model for the analysis and design of organizations in multi-agent systems. In *Proceedings of Third International Conference on Multi-Agent Systems (ICMAS'98)*, IEEE Computer Society, 128-135
- Ferber, J., Michel, F., Baez-Barranco, J.-A. (2004). AGRE: Integrating Environments with Organizations. In *Proceedings of E4MAS*, 48-56
- Fox M, Barbuceanu M, Gruninger M, Lin J (1997). An Organization Ontology for Enterprise Modelling, in *Simulating Organizations: Computational Models of Institutions and Groups*, edited by M. Prietula, K. Carley and L. Gasser, Menlo Park CA: AAAI/MIT Press, 131-152

- Franklin, S., Graesser, A. (1996) Is it an agent, or just a program? A taxonomy for autonomous agents. In Mueller, J.P., Wooldridge, M., and Jennings, N.R., editors, *Intelligent Agents III*, pages 21-35. Springer LNAI 1193
- Galbraith JR (1978). *Organization design*, Addison-Wesley Publishing Company, London Amsterdam Sydney
- GERAM (2003). GERAM: The Generalized Enterprise Reference Architecture and Methodology. IFIP-IFAC Task Force on Architectures for Enterprise Integration, In: Bernus P, Nemes L, Schmidt G. (Eds): *Handbook on Enterprise Architectures*, Springer-Verlag, 21-63
- Gertman DI, Blackman HS (1994). *Human reliability and safety analysis data handbook*. John Wiley & Sons, Inc., New York, USA
- Giddens A (2006) *Sociology*, 5th edn, Polity, Cambridge
- Glendon AI, Stanton NA (2000). Perspectives on safety culture. *Safety Science* 34:193-214
- Goeters KM (2004). *Aviation psychology: practice and research*. Ashgate, Aldershot, UK
- Gordon R, Kennedy R, Mearns K, Jensen CL, Kirwan B (2006). *Understanding safety culture in air traffic management*. Eurocontrol, report EEC Note No. 11/06
- Groeneweg J. *Controlling the controllable: the management of safety*. DSWO Press, Leiden University, The Netherlands, 4th edition, 1998
- Gronau, N. (2004) *Enterprise Resource Planning und Supply Chain Management - Architektur und Funktionen*. Oldenbourg Wissenschaftsverlag, Munchen
- Grossi, D., Aldewereld, H., Vazquez-Salceda, J., Dignum, F. (2006) Ontological Aspects of the Implementation of Norms in Agent-Based Electronic Institutions. *Journal of Computational and Mathematical Organization Theory*. Special issue of Normative Multiagent Systems. Springer 2006.12 (2-3), pp. 251-275
- Grote G, Helmreich RL, Sträter O, Häusler R, Zala-Mezö E, Sexton JB (2004). Setting the stage: Characteristics of organizations, teams and tasks influencing team processes. In: Dietrich R, Childress TM. *Group interaction in high risk environment*. Ashgate, Aldershot, England
- Guldenmund FW (2000). The nature of safety culture: a review of theory and research. *Safety Science* 34:215-257
- Gurnell, D. (2006). *A cooperative study of approaches to multi agent planning*. PhD thesis.
- Hackman JR (1980) Work redesign and motivation. *Professional Psychology*, 11: 445-455
- Hale AR (2000). Editorial: Culture's confusions. *Safety Science* 34:1-14
- Hannan MT, Freeman J (1977) The population ecology of organizations. *American Journal of Sociology* 82: 929-964
- Hannoun, M., Boissier, O., Sichman JS., Sayettat, C. (2000). MOISE: An Organizational Model for Multi-agent Systems. In *Proceedings of the International Joint Conference, 7th Ibero-*

- American Conference on AI: Advances in Artificial Intelligence, LNCS*, vol. 1952, 156 – 165
- Hawkins FH (1987). *Human factors in flight*. Ashgate, Aldershot, England
- Henesey, L., Wernstedt, F., Davidsson, P., (2003) Market-Driven Control in Container Terminal Management. *Proceedings of the 2nd International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT'03)*, Hamburg, Germany
- Hodgson, A., Roennquist, R., Busetta, P. and Howden, N. (2000) Team Oriented Programming with SimpleTeam, in *Proc. of SimTecT 2000*, Sydney, Australia, 2000, pp. 115-122
- Hollnagel E (1998). *Cognitive reliability and error analysis method: CREAM*. Elsevier Science Ltd, Oxford, England
- Hollnagel E (2004). *Barriers and accident prevention*. Ashgate, Aldershot, England
- Hollnagel E, Woods DD (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press, Boca Raton (FL), USA
- Hollnagel E, Woods DD, Leveson N (eds.) (2006). *Resilience engineering: Concepts and precepts*. Ashgate, Aldershot, England
- Hopkins A (2006). Studying organizational cultures and their effects on safety. *Safety Science* 44:875-889
- Hoogendoorn, M., Jonker, C.M., Popova, V., and Sharpanskykh, A. (2006) Automated Verification of Disaster Plans in Incident Management. *Disaster Prevention and Management* (in press).
- Horling B, Lesser V (2005). A Survey of multi-agent organizational paradigms. *The Knowledge Engineering Review*, 19(4): 281-316
- Hudson PTW, Reason JT, Wagenaar WA, Bentley PD, Primrose M, Visser JP. Tripod Delta: Proactive approach to enhanced safety. *Journal of Petroleum Technology* 46:58-62, 1994
- Huhns, MN, Stephens, LM. (1999) Multiagent Systems and Societies of Agents, in *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, Gerhard Weiss, editor, MIT Press, Cambridge, MA
- ICAO (2002). *Human factors guidelines for safety audits manual*. Doc 9806 AN/763
- Ingrand, FF, Georgeff, MP, Rao, AS (1992) An Architecture for Real-time Reasoning and System Control, *IEEE Expert* 7(6): 33-44
- Isaac A, Shorrock ST, Kenndy R, Kirwan B, Andersen H, Bove T (2003). *The human error in ATM technique (HERA-JANUS)*. Eurocontrol, report HRS/HSP-002-REP-03
- Isaac A, Straeter O, Van Damme D (2004). *A method for predicting human error in ATM (HERA-PREDICT)*. Eurocontrol, report HRS/HSP-002-REP-07
- Jakobson, G., Parameswaran, N., Buford, J., Lewis L., and Ray P. (2006) Situation-Aware Multi-Agent System for Disaster Relief Operations Management. In *Proceedings of the Third International Conference on Information Systems for Crisis Response and Management ISCRAM 2006*, 313-326

- Jennings, NR (1994). *Cooperation in industrial multi-agent systems*. Number 43 in Computer Science. World Scientific
- Jennings, N. R., Corera, J. and Laresgoiti, I. (1995) Developing Industrial Multi-Agent Systems (Invited Paper). In *Proceedings of 1st Int. Conf. on Multi-Agent Systems (ICMAS '95)*, pp. 423-430, San Francisco, USA
- Jonker CM, Sharpanskykh A, Treur J, Yolum P (2006), Design Operators to Support Organizational Design. In: J. Gero (ed.), *Proceedings of the Second International Conference on Design Computing and Cognition, DCC-06*. Springer Verlag, 203-222
- Jonker C.M., Sharpanskykh, A., Treur, J. and Yolum, P. (2007) A Framework for Formal Modeling and Analysis of Organizations, *Applied Intelligence* (in press)
- Kast FE, Rosenzweig JE (1972) General systems theory: applications for organization and management. *Academy of Management Journal* December: 447-465
- Kennedy R, Kirwan B (1998). Development of a hazard and operability-based method for identifying safety management vulnerabilities in high risk systems. *Safety Science* 30: 249-274
- Kirwan B (1994). *A guide to practical human reliability assessment*. Taylor & Francis Ltd, London, UK
- Kirwan B, Kennedy R, Taylor-Adams S, Lambert B (1997). The validation of three Human Reliability Quantification techniques – THERP, HEART and JHEDI: Part II – Results of validation exercise. *Applied Ergonomics* 28(1):17-25
- Kirwan B, Gibson H, Kennedy R, Edmunds J, Cooksley G, Umbers I (2004). Nuclear Action Reliability Assessment (NARA): A data-based HRA tool. In: Spitzer C, Schmocker U, Dang VN (eds.), *Probabilistic Safety Assessment and Management, PSAM 7 - ESREL '04*, Springer, London
- Langlois, R.N., Robertson, P. L. (1995) *Firms, Markets, and Economic Change: A Dynamic Theory of Business Institutions*. London: Routledge
- Le Coze J (2005). Are organizations too complex to be integrated in technical risk assessment and current safety auditing? *Safety Science* 43:613-638
- Lee TR (1996). Perceptions, attitudes and behaviour: the vital elements of a safety culture. *Health and Safety*, October 1:15
- Leveson N (2004). A new accident model for engineering safer systems. *Safety Science* 42:237-270
- Leveson NG, Barrett B, Carroll J, Cutcher-Gershenfeld J, Dulac N, Zipkin D (2005). *Modeling, analyzing and engineering NASA's safety culture*. Phase 1 Final report
- Ljungberg, M. (1992) *The Oasis Air Traffic Management System*, Tech. Note 28, Australian Artificial Intelligence Institute, Carlton, Australia
- Lorsch JW, Lawrence PR (1970). *Organization design*, Richard D. Irwin Inc., USA



- Luxhoj JT (2003). Probabilistic causal analysis for system safety risk assessments in commercial air transport. *Proceedings of the Workshop on Investigating and Reporting of Incidents and Accidents (IRIA)*, September 16-19, 2003, Williamsburg (VA)
- Marais K, Saleh JH, Leveson NG (2006). Archetypes for organizational safety. *Safety Science* 44:565-582
- March, JG, Simon, HA (1967) *Organizations*, John Wiley & Sons, Inc.
- Mellett U, Nendick M (2007). *The Human Factors Case: Managing human factors issues for ATM projects*. Eurocontrol report, edition 2.0, 15 March 2007
- Menzel, C., Mayer, R.J.: The IDEF family of languages. In: Bernus, P. et al. (eds.): *Handbook on Architectures of Information Systems*, Springer-Verlag, Heidelberg (1998) 209-241
- Mintzberg H (1979) *The Structuring of Organizations*, Prentice Hall, Englewood Cliffs
- Montijn C, De Jong H (2006). *Safety culture in air transport: Definition, characteristics, indicators and classification scheme*. National Aerospace Laboratory NLR, memorandum ATSF-2006-150
- Mooney JD (1947) *The principles of organization*, Harper & Bros., New York
- Morgan G (1996). *Images of organizations*, SAGE Publications, Thousand Oaks London New Delhi
- Omicini, A. (2000). SODA: Societies and infrastructures in the analysis and design of agent-based systems. In *Proceeding of AOSE'00*, 185–193
- Pariès J (2006). Complexity, emergence, resilience... In: Hollnagel E, Woods DD, Leveson N (eds.). *Resilience engineering: Concepts and precepts*. Ashgate, Aldershot, England
- Parker D, Lawrie M, Hudson P (2006). A framework for understanding the development of organizational safety culture. *Safety Science* 44:551-562
- Perrow C (1984). *Normal accidents: Living with high-risk technologies*. Basic Books, New York, USA
- Pfeffer J (1982). *Organizations and organization theory*, Pitman Books Limited, Boston London Melbourne Toronto
- Pfeffer J, Salancik GR (1978). *The external control of organizations: A resource dependence perspective*, Harper & Row, New York
- Pidgeon NF (1991). Safety culture and risk management in organizations. *Journal of Cross-Cultural Psychology* 22(1):129-140
- Pidgeon N, O'Leary M (2000). Man-made disasters: why technology and organizations (sometimes) fail. *Safety Science* 34:15-30
- Pitt, J., Kamara, L., Sergot, MJ, Artikis, A (2006) Voting in Multi-Agent Systems. *Comput. J.* 49(2): 156-170
- Popova V, Sharpanskykh A (2007a). Formal Modelling of Goals in Agent Organizations. In *Proceedings of Agent Organizations: Modelling and Simulation Workshop during the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*

- Popova V, Sharpanskykh A (2007b). Modeling Organizational Performance Indicators. In: *Proceedings of International Modeling and Simulation Multiconference, invited session on Agent Based Modeling and Simulation, in Industry and Environment*, SCS Press
- Popova V, Sharpanskykh A (2007c). Process-Oriented Organization Modeling and Analysis. In: *Proceedings of the 5th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems (MSVVEIS 2007)*, INSTICC Press
- Popova V, Sharpanskykh A (2007d). Formal Analysis Of Executions Of Organizational Scenarios Based On Process-Oriented Models. In: *Proceedings of 21st EUROPEAN Conference on Modelling and Simulation ECMS 2007*, SCS Press
- Popova V, Sharpanskykh A (2007e). A Formal Framework for Modeling and Analysis of Organizations. In: *Proceedings of the Situational Method Engineering Conference, ME'07*, Springer Verlag
- Rasmussen J, Svedung I (2000). *Proactive risk management in a dynamic society*. Swedish Rescue Services Agency
- Reason J (1990). *Human error*. Cambridge University Press, Cambridge, UK
- Reason J (1997). *Managing the risk of organizational accidents*. Ashgate, Aldershot, England
- Reason J, Hollnagel E, Paries J (2006). *Revisiting the Swiss cheese model of accidents*. Eurocontrol, EEC Note no. 13/06
- Regis, V, Bryan, H., Lesser, V. (2000) Experiences in Simulating Multi-Agent Systems Using TAEMS. *The Fourth International Conference on MultiAgent Systems (ICMAS 2000)*, AAAI
- Reiman T, Oedewald P (2007). Assessment of complex sociotechnical systems – Theoretical issues concerning the use of organizational culture and organizational core task concepts. *Safety Science* 45(7):745-768
- Romme, AGL (2003). Making a difference: Organization as design. *Organization Science*, 14, 558-573
- Rosenschein, JS. and Zlotkin, G. (1994). *Rules of encounter: Designing convention for automated negotiation among computers*. Artificial Intelligence. The MIT Press.
- Rosqvist T (2003). *On the use of expert judgement in the qualification of risk assessment*. Espoo 2003. Technical Research Centre of Finland, VTT Publications 507
- Russell S, Norvig P. (1995) *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Inc.
- Scarborough A, Bailey L, Pounds J (2005). *Examining ATC operational errors using the human factors analysis and classification system*. Federal Aviation Administration, report DOT/FAA/AM-05/25
- Scheer A-W, Nuettgens M (2000). ARIS Architecture and Reference Models for Business Process Management. In: van der Aalst, W.M.P.; Desel, J.; Oberweis, A. (eds.), *LNCS 1806*, Berlin et al. 366-389
- Schein EH. *Organizational culture and leadership*. Jossey-Bass, San Francisco, USA

- Schenk D. and Wilson P. (1994) *Information Modeling: The EXPRESS Way*. Oxford University Press
- Schillo, M., Spresny, D. (2005) Organization: The Central Concept for Qualitative and Quantitative Scalability, in *Socionics: Contributions to the Scalability of Complex Social Systems* edited by K. Fischer and M. Florian, Lecture Notes in Artificial Intelligence, Berlin, vol. 3413, Springer, pp. 84-103
- Schoenharl T, Bravo R, Madey G (2006) WIPER: Leveraging the Cell Phone Network for Emergency Response. In *The International Journal of Intelligent Control and Systems*, vol. 11(4): 209- 216
- Schoemaker, P. (1982) The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. In *Journal of Economic Literature*, vol. 20:529-563
- Scott WG, Mitchell TR, Birnbarum PH (1981) *Organization theory: a structural and behavioural analysis*, Richard D. Irwin inc., Illinois, USA
- Sharpanskykh, A. (2007) Authority and its Implementation in Enterprise Information Systems. In *Proceeding of the 1st International Workshop on Management of Enterprise Information Systems, MEIS 2007*, INSTICC Press
- Shorrock ST (2002). The two-fold path to human error analysis: TRACER-lite retrospection and prediction. *Safety Systems (Newsletter of the Safety-Critical Systems Club)*, vol. 11, no. 3
- Shorrock ST, Kirwan B (2002). Development and application of a human error identification tool for air traffic control. *Applied Ergonomics* 33:319-336
- Shorrock S, Woldring M, Hughes G (2004). *The human factors case: Guidance for human factors integration*. Eurocontrol, report HRS/HSP-003-GUI-01, August 2004
- Sichman, JS, Conte, R., Castelfranchi, C., and Demazeau, Y. (1994) A Social Reasoning Mechanism Based on Dependence Networks. In A G. Cohn (Ed.), *Proceedings of the 11th. European Conference on Artificial Intelligence*, Baffins Lane, England: John Wiley & Sons
- Simon, Herbert (1957). A Behavioral Model of Rational Choice, in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley
- Simona K, Mengolini A, Bolado-Lavin R (2005). *Formal expert judgement: an overview*. European Commission Joint Research Centre, report EUR 21772 EN
- Singh MP (1996) Synthesizing distributed constrained events from transactional workflow specifications. In *Proceedings of the 12th IEEE International. Conference on Data Engineering*, 616–623
- Spouge J, Perrin E (2006). *Main report for the 2005/2012 Integrated Risk Picture for air traffic management in Europe*. Eurocontrol, EEC Note no. 05/06
- Sträter O (2005). *Cognition and safety: An integrated approach to system design and assessment*. Ashgate, Aldershot, England



- Stroeve SH, Blom HAP, Van der Park MNJ (2003). Multi-agent situation awareness error evolution in accident risk modelling. *Proceedings 5th USA/Europe Air Traffic Management R&D Seminar*, Budapest, Hungary
- Stroeve SH, Blom HAP, Bakker GJ (2006). Safety risk impact analysis of an ATC runway incursion alert system. *Eurocontrol Safety R&D Seminar*, Barcelona, Spain
- Swain AD (1990). Human reliability analysis: Need, status, trends and limitation. *Reliability Engineering and System Safety* 29:301-313
- Swain AD, Guttman HE (1983). *A handbook of human reliability analysis with emphasis on nuclear power plant applications*. USNRC-Nureg/CR-1278, Washington DC-20555
- Sycara, K. (1998) Multiagent Systems. *Artificial Intelligence*, 19(2):79–92
- Tambe, M. (1997) Agent Architectures for flexible, practical teamwork. In *Proceedings of the AAAI American Association of Artificial Intelligence*
- Tatomir, B., Rothkrantz, L. and Popa, M. (2006) Intelligent system for exploring dynamic crisis environments. In *Proceedings of the Third International Conference on Information Systems for Crisis Response and Management ISCRAM 2006*, 288-297
- Tessier C, Müller HJ, Fiorino H, Chaudron L (2001). Agents' conflicts: new issues. In Tessier C, Chaudron L, Müller HJ, *Conflicting agents: conflict management in multi-agent systems*. Kluwer Academic Publishers, Norwell, USA
- Turner BA (1978). *Man-made disasters*. Wykeham Science Press, London, UK
- Uttal B (1983). The corporate culture vultures. *Fortune*, Oct. 17: 66-72
- Van der Aalst, W. M.P., van Hee, K (2002) *Workflow Management: Models, Methods, and Systems*. MIT Press
- Van der Hoek, W., van Linder B., and Ch. Meyer, J.-J. (1998) An Integrated Modal Approach to Rational Agents, in: M. Wooldridge and A. Rao (eds.), *Foundations of Rational Agency, Applied Logic Series* 14, Kluwer, Dordrecht, pp. 133–168
- Von Thaden TL, Wiegmann DA, Shappell SA (2006). Organizational factors in commercial aviation accidents. *The International Journal of Aviation Psychology* 16(3):239-261
- Vroom VH (1964). *Work and motivation*. Wiley, New York
- Walter B (1968) *Modern systems research for the behavioral scientist*, Aldine Publishing Co, Chicago
- Warner, M., Witzel, M. (2004) *Managing in Virtual Organizations*, Thomson Learning
- Wickens CD, Hollands JG (1999). *Engineering psychology and human performance*. Prentice Hall, 1999
- Wiegmann DA, Shappell SA (2001). Human error analysis of aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Aviation, Space, and Environmental Medicine* 72(11):1006-1016
- Weiss, G. (ed.) (1999) *Multiagent Systems, A Modern Approach to Distributed Artificial Intelligence*, MIT Press, MA



- Wilkins, DE, Meyers, KL (1995) A common knowledge representation for plan generation and reactive execution. *Journal of Logic and Computation*, 5(6): 731-761.
- Williams JC. A data-based method for assessing and reducing human error to improve operational performance. *Proceedings IEEE Conference on Human Factors in Power Plants*, Monterey (CA), 1988
- Wooldridge, MJ (2000) *Reasoning about Rational Agents*. The MIT Press, Cambridge, MA.
- Yeatts DE, Hyten C (1998). *High-performing self-managed work teams: A comparison of theory to practice*. Sage, Thousand Oaks
- Yu, E. (1997). Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering. *3rd IEEE Int. Symposium on Requirements Engineering*, 226-235
- Yukl G (2006). *Leadership in organizations*, 6edn, Englewood Cliffs, NJ: Prentice-Hall
- Zambonelli, F., Jennings, N. R., Wooldridge, M. (2003). Developing multiagent systems: the Gaia Methodology, *ACM Transactions on Software Engineering and Methodology*, vol. 12 (3): 317-370

Appendix A Summaries of risk assessment methods

A.1 HFACS

The Human Factors Analysis and Classification Systems (HFACS) is a tool for analysis of human contributions to aviation accidents (Wiegmann and Shappell, 2001). Drawing on the human error concept of Reason (1997), which describes active as well as latent failures, HFACS describes four main levels of failures:

1. unsafe acts of operators,
2. preconditions for unsafe acts,
3. unsafe supervision, and
4. organizational influences.

Each of these levels contains several subcategories shown in Figure 19.

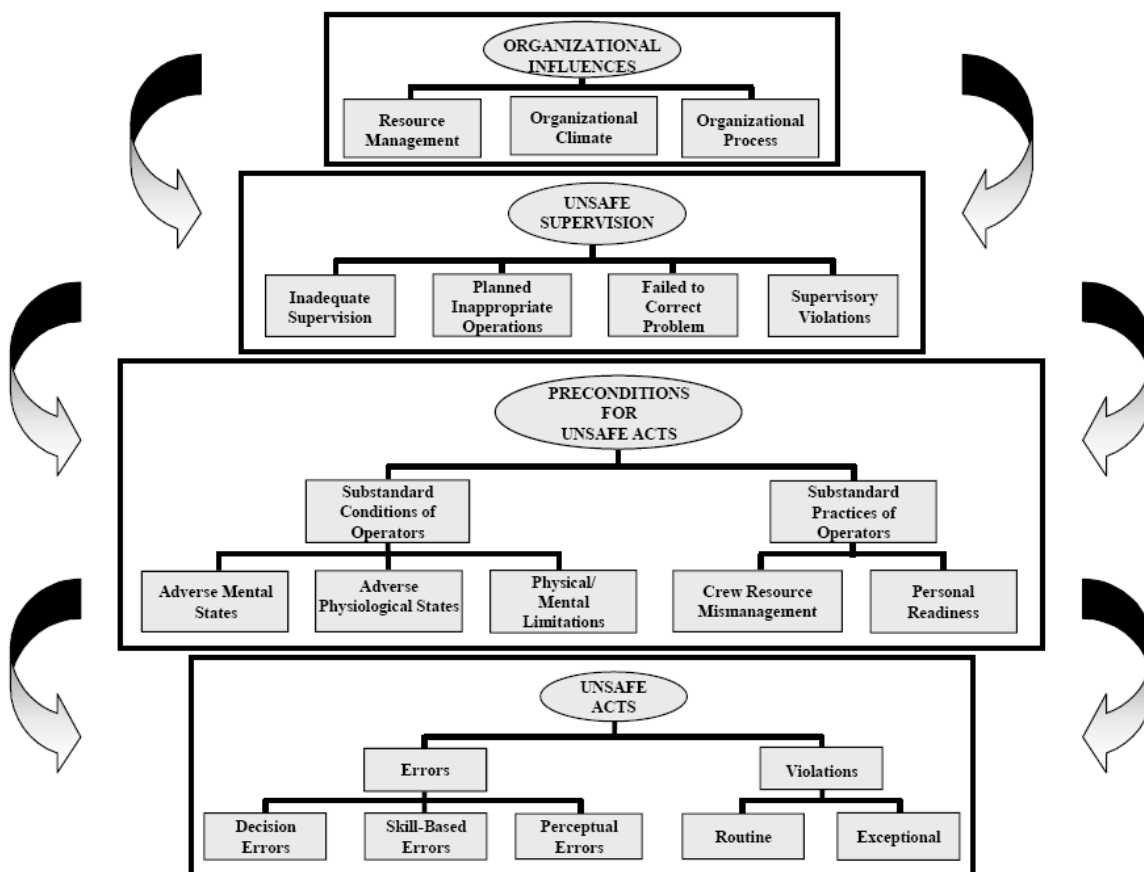


Figure 19: Overview of the Human Factors Analysis and Classification System (HFACS) (Wiegmann and Shappell, 2001)



The HFACS taxonomy includes specific actions / omissions / conditions for each of these subcategories. Examples for some of the subcategories of Figure 19 are:

- *Skill-based errors*: breakdown in visual scan, failed to prioritise attention, omitted step in procedure, poor technique, over-controlled the aircraft, etc.;
- *Decision errors*: improper procedure, misdiagnosed emergency, wrong response to emergency, exceed ability, etc.;
- *Adverse mental states*: channel attention, complacency, haste, loss of situational awareness;
- *Crew resource mismanagement*: failed to communicate/coordinate, failure of leadership, failed to conduct adequate brief;
- *Inadequate supervision*: failed to provide guidance, failed to provide training, failed to track performance;
- *Supervisory violations*: authorised unnecessary hazard, failed to enforce rules and regulations;
- *Resource management*: selection, training, lack of funding, staffing;
- *Organizational climate*: norms and rules, hiring and firing, organizational justice.

The HFACS categorisation provides the basis for a database of accident contributory / causal factors. HFACS analysis has been used to identify human factors problems in aviation and military accidents (Wiegmann and Shappell 2001). HFACS data has been used in the development of a Bayesian Belief Network for aviation risk (Aviation System Risk Model; Luxhoj 2003).

HFACS was used for the analysis of organizational factors in 60 U.S. commercial aviation accidents in the period 1990 to 2000 (Von Thaden et al. 2006), including 17 accidents of major airlines (FAR Part 121) and 43 accidents of feeder airlines and commuter and on-demand operators (FAR Part 135). The organizational factors that were identified to contribute highest to the accident causation are:

- Inadequate procedures or directives (21%): ill-defined or conflicting policies;
- Inadequate training (18%): opportunities for initial, upgrade or emergency training of pilots not implemented or made available to pilots;
- Inadequate surveillance of operations (13%): organizational climate issues, chain-of-command, quality assurance and trend information;
- Insufficient standard/requirements (12%): clearly defined organizational objectives, adherence to policy;
- Inadequate information sharing (12%): logbooks, updates, weather reports;
- Inadequate supervision of operations (10%): failure to provide guidance, oversight and leadership to operations;
- Company/management induced pressure (6%): threats to pilot job status and/or pay;
- Faulty documentation (4%): inaccurate checklists, signoffs, company records;



- Inadequate substantiation process (3%): accountability, regulation, recording/reporting process;
- Inadequate facilities (1.5%): environmental controls, lighting, etc for flight operations.

The data show differences in the types of organizational contributory factors for large versus small airlines. Inadequacies in procedures and directives rank among the highest organizational problems for both large and small airlines alike. For small airlines, training, surveillance, and supervision also tend to be significant problems. As airlines grow larger, organizational problems appear to shift from issues of training and surveillance to issues of information sharing, communication, and documentation.

A.2 Tracer/HERA

TRACER (Technique for the Retrospective and Predictive Analysis of Cognitive Errors) is a technique for the retrospective and prospective analysis of human errors in air traffic control (Shorrock and Kirwan 2002). A reduced, easier to use version of has been published as TRACER-lite (Shorrock 2002). TRACER was used by EUROCONTROL as a basis to develop similar tools for retrospective analysis (HERA-JANUS, Isaac et al. 2003) and prospective analysis (HERA-PREDICT; Isaac et al. 2004) of human error in ATM.

TRACER uses the following taxonomy for the origin of cognitive errors (the terminology used in HERA is between brackets):

- External error modes (HERA: error type): the external and observable manifestation of the error;
- Cognitive domain (HERA: error detail): the cognitive processes involved in the error;
- Internal error modes (HERA: error mechanism): type and manner of cognitive function failure;
- Psychological error mechanism (HERA: information processing levels): psychological explanation of the internal error mode.

Table 8 shows aspects considered in this taxonomy and examples of the errors at the various levels.

The context in which the error arises is described at the level of tasks performed when the error happened, used information and equipment, and other general contextual conditions or performance shaping factors. Table 9 shows keywords of the context description.

Table 8: Error classification in HERA-JANUS technique (Isaac et al. 2003).

Levels	Examples
Error Type: External manifestation of the error	
Timing of action	<ul style="list-style-type: none"> • Action too early • Right action in the wrong order
Selection of action	<ul style="list-style-type: none"> • Wrong action



	<ul style="list-style-type: none"> • Right action to the wrong aircraft
Information transfer	<ul style="list-style-type: none"> • Transmitted/sent incorrect information • Failed to get required information
Error / Rule breaking and violation	<ul style="list-style-type: none"> • Intended violation • Possible reckless violation • Possible negligent rule breaking • Possible organizational induced violation • Routine rule breaking
Error detail: Cognitive process involved in the error	
Perception and vigilance	Visual, auditory detection
Response execution	Speech, motor control
Memory	Short term, long term
Planning and decision making	Prediction, planning
Error mechanism: Type and manner of cognitive function failure	
Perception and vigilance	<ul style="list-style-type: none"> • Hearback error • Late auditory recognition • No detection (visual)
Response execution	<ul style="list-style-type: none"> • Selection error • Incorrect information transmitted
Working memory	<ul style="list-style-type: none"> • Forget to monitor • Forget planned action
Long-term memory	<ul style="list-style-type: none"> • No recall of information • Misrecall information
Planning and decision making	<ul style="list-style-type: none"> • Misperjection of aircraft • No decision or plan • Insufficient plan
Information processing levels: Psychological mechanism of error	
Perception and vigilance	<ul style="list-style-type: none"> • Visual search failure • Expectation bias • Spatial confusion • Discrimination failure
Response execution	<ul style="list-style-type: none"> • Unclear speech • Intrusion of thoughts • Slip of the tongue
Working memory	<ul style="list-style-type: none"> • Memory capacity overload • Distraction • Preoccupation
Long-term memory	<ul style="list-style-type: none"> • Insufficient learning • Rarely used information
Planning and decision making	<ul style="list-style-type: none"> • Denied risk • Incorrect prioritisation • Incorrect knowledge

Table 9: Context description in HERA-JANUS technique (Isaac et al. 2003).

Tasks while error occurred	Information & Equipment	Contextual conditions
Coordination	Controller materials	Pilot-controller communications
Tower observation	Controller activities	Pilot actions
Planning	Variable aircraft information	Traffic and airspace
R/T communication and instruction	Fixed aircraft information	Weather
Control room communications	Time and location	Documentation and procedures
Strip work	Airport	Training and experience
Materials checking	Flight rules	Workplace design and HMI
Radar monitoring	Communication	Environment
HMI input and functions	Navigation	Personal factors
Handover briefing	Surveillance	Team factors
Takeover	Visual approach aids	Organizational factors
Training	Aerodrome equipment warning panels	
Supervision	Aerodrome auxiliary equipment controls	
Check / Examination	Flight information displays	
	Input devices	
	Other information displays	

The cognitive error taxonomy is used both in TRACER and in HERA for retrospective and prospective error analysis. This two-sided perspective of analysis processes is illustrated in Figure 20 for TRACER (Shorrock and Kirwan 2002). In doing a retrospective analysis, the analyst is supported by a number of flowcharts that guide the classification process in both TRACER (Shorrock and Kirwan, 2002) and HERA (Isaac et al. 2003). The process for a prospective analysis in TRACER (Shorrock and Kirwan 2002) and HERA-PREDICT (Isaac et al. 2004) is based on a task analysis and subsequent identification and classification of potential errors using the error taxonomy. Moreover, in HERA-PREDICT as a last step a (simple) risk assessment of the task is made.

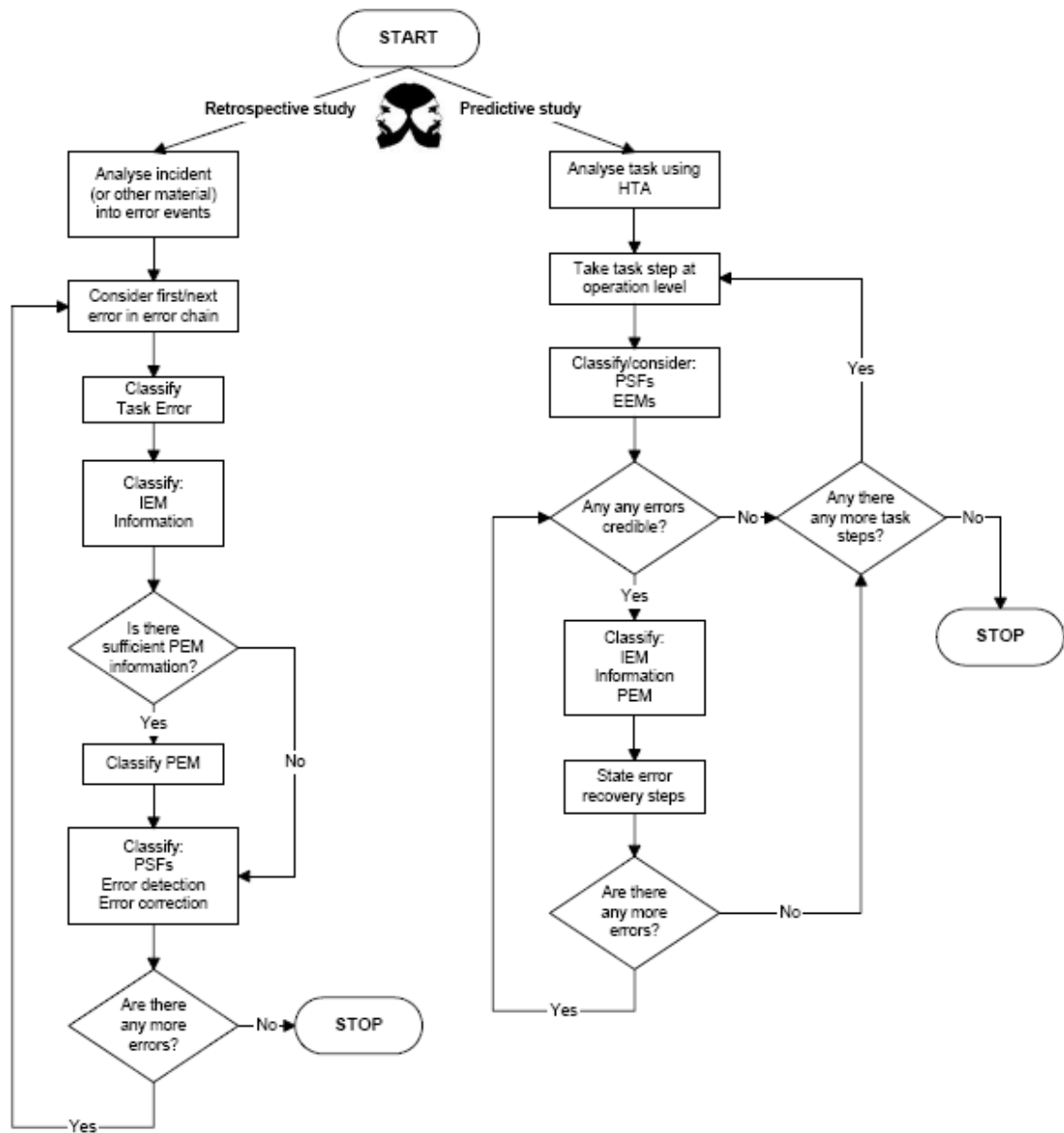


Figure 20: Processes of using TRACER taxonomy for retrospective and prospective analyses (two-sided Janus perspective) (Shorrock and Kirwan 2002).

A.3 HEART

The Human Error Assessment and Reduction Technique (HEART) includes a relatively simple HEP quantification process with a limited number of steps. It is based on an extensive review of human-performance literature. It works by using a set of Generic Task Types (GTTs) for tasks in an HRA (Williams 1988); Kirwan 1994). Each GTT has a nominal HEP and confidence bound. The effect of contextual conditions is represented by factors for Error Producing Conditions (EPCs), which increase the error probability (i.e., with values >1), and factors for Assessed Proportion Of Affect (APOA), which describe the relative contribution of the EPC for



the situation considered (i.e., with values between 0 and 1). Now the HEP of a GTT given a specific contextual condition is:

$$P_{HE} = P_{HE}^{GTT} \prod_i [(f_i^{EPC} - 1)c_i^{APOA} + 1].$$

HEART provides values for P_{HE}^{GTT} and f_i^{EPC} . Selection of appropriate GTTs (8 types are defined in HEART, see Table 10), EPCs (26 conditions are defined in HEART, see top list in Table 11) and values for c_i^{APOA} is up to the assessor.

Table 10: HEART error probabilities for generic task types (Kirwan, 1994). Note that these values are not based on aviation data.

Generic task type (GTT)	Error probability P_{HE}^{GTT}	
	Nominal	5 - 95% bounds
Totally unfamiliar, performed at speed with no real idea of consequences	$5.5 \cdot 10^{-1}$	$3.5 \cdot 10^{-1} - 9.7 \cdot 10^{-1}$
Shift or restore system to a new or original state on a single attempt without supervision or procedure	$2.6 \cdot 10^{-1}$	$1.4 \cdot 10^{-1} - 4.2 \cdot 10^{-1}$
Complex task requiring high level of comprehension and skill	$1.6 \cdot 10^{-1}$	$1.2 \cdot 10^{-1} - 2.8 \cdot 10^{-1}$
Fairly simple task performed rapidly or given scant attention	$9 \cdot 10^{-2}$	$6 \cdot 10^{-2} - 1.3 \cdot 10^{-1}$
Routine, high-practiced, rapid task involving relatively low level of skill	$2 \cdot 10^{-2}$	$7 \cdot 10^{-3} - 4.5 \cdot 10^{-2}$
Restore or shift a system to original or new state following procedures, with some checking	$3 \cdot 10^{-3}$	$8 \cdot 10^{-4} - 7 \cdot 10^{-3}$
Completely familiar, well designed, highly practiced routine task occurring several times per hour	$4 \cdot 10^{-4}$	$8 \cdot 10^{-5} - 9 \cdot 10^{-3}$
Respond correctly to system command when there is an augmented or automated supervisory system	$2 \cdot 10^{-5}$	$6 \cdot 10^{-6} - 9 \cdot 10^{-4}$

Table 11: Most prominent HEART error producing conditions (Kirwan, 1994) (total list contains 26 EPCs). Note that these values are not based on aviation data.

Error producing condition (EPCs)	Maximum factor f_i^{EPC}
Unfamiliarity	17
Shortage of time	11
Low signal to noise ratio	10
Ease of information suppression	9
Ease of information assimilation	8
Model mismatch: operator / designer	8
Difficulty reversing an unintended action	8

Channel capacity overload	6
Technique unlearning and apply one with opposing philosophy	6
Et cetera	et cetera

HEART has been the principal tool for HRA in the nuclear industry for the last decade in the UK (Kirwan et al. 2004). A validation exercise for three HRA techniques (HEART, THERP, JHEDI) shows that the overall precision of the estimated HEPs is within a factor 10 of the true HEPs for 72% of the cases and it is within a factor 3 for 38% of the cases (Kirwan et al. 1997). Similar performance was found for the three techniques evaluated.

HEART has been used by NATS for human failures quantification of events in fault tree modelling of ATC operations for two airspace sectors in the UK (CAA 1993). EUROCONTROL is using HEART in combination with Absolute Probability Judgement (APJ) and Paired Comparisons (PC) for human failure quantification.

A.4 Absolute probability judgement

Expert judgement is used extensively in risk assessments. If the term ‘expert’ is used in a wide sense to include normative experts (risk assessors) (Rosqvist 2003), expert judgement is applied in all risk assessments; used in the more regular fashion of a domain expert (used in the remainder of this section), expert judgement is still applied in lots of risk assessments. Expert judgement can be used in various steps of a risk assessment cycle (De Jong et al. 2006), e.g. hazard identification, risk (severity/frequency) evaluation, identification of risk mitigating measures. In the current context, we consider qualitative or quantitative evaluation of severity/frequency in risk scenarios with emphasis on quantitative estimates of model parameters, such as event probabilities assessed in Absolute Probability Judgement (APJ).

Expert judgement techniques, whether qualitative or quantitative in nature, seek to avoid biases in expert judgement. There are a number of well-documented biases (Simona 2005), including

- Availability biases: giving more weight to recent or otherwise memorable events;
- Conservatism biases: underestimating extremes such as very high and very low probabilities or frequencies;
- Anchoring biases: inadvertently giving the expert a ‘clue’ as to the ‘desired’ number, hence making it difficult for them to come up with a highly different number, despite what they originally thought;
- Motivational biases: one or more experts have some vested interest (known or unknown to themselves) in deriving a particular answer – e.g. a designer quantifying the failure likelihood of his or her own design.
- Group dynamic biases (in case of expert judgement group techniques): group dynamics in expert discussions may lead to biases, e.g. one or more experts may dominate a discussion.

For Absolute Probability Judgement (APJ), there are single expert and group methods. Group methods can be discerned in four types (Kirwan et al. 1994):

- Aggregated individual methods: experts make probability estimates individually and the judgements are aggregated;
- Delphi method: experts make their estimates individually, next review each others' assessments, then reassess their judgements, and the judgements are aggregated;
- Nominal group technique: similar to the Delphi method, except that the allowed discussion between experts is limited to the clarification of comments;
- Consensus group method: a group of experts discusses together to find an estimate upon which all group members agree.

Some recent studies (Cooke and Goossens 2000; Rosqvist 2003; Simona et al. 2005) focus on aggregated individual methods for 'formal' expert judgement of a distribution of a parameter value (primarily event probabilities) and a weighting process for the aggregation of expert judgements. An example of application of formal expert judgement techniques in air traffic is in the ongoing research on development of a causal model for air traffic (Ale et al., 2006).

A.5 Use of incident/accident databases

Accident and incident databases can be used in support of the evaluation of human performance in the context of safety assessments. A database is considered a tool to collect and store data for analysis, rather than a methodology in itself.

Many organizations in the aviation industry have systems in place for reporting safety related events, which are stored in accident and incident databases. The events stored consider accidents, incidents and occurrences and may address a variety of issues, including flight operational issues, maintenance occurrences, air traffic control related events and aircraft system failures. Records may contain information on date and location of the event, aircraft type involved, flight phase, narrative on what happened, classification of causal factors present in the event, information on fatalities and injuries etc. Information on human or organizational factors may be present in the database records depending on what the reporter relayed. Databases use different structures and taxonomies to store events, and classify human and organizational factors. Commercial off-the-shelf tools are available as database tool.

Many national civil aviation authorities and accident investigation boards have database to collect accident and incidents, usually on events which have occurred in that country or with carriers operating from/in that country. Examples are the United States National Transportation Safety Board (NTSB) accident database, Federal Aviation Administration Accident and Incident database, and Flight Operational Quality Assurance (FOQA) programme. Some reporting systems and associated databases are specifically aimed at collecting mandatory occurrence

reports, and some databases are voluntary reporting systems, e.g. the Aviation Safety Reporting System (ASRS) in the United States, Confidential Human Factors Incident Reporting Programme (CHIRP) in UK or Confidential Aviation Incident Reporting (CAIR)/Aviation Self-Reporting Scheme (ASRS) in Australia, which are publicly accessible for submitting reports and also frequently report results of queries in magazines or allow ‘public queries’ in the database. In addition, most airlines have databases to store airline safety reports submitted by flight crews, cabin crews and other personnel on safety related and operational issues in the day-to-day flight operations. A prominent example is the British Airways Safety Information System (BASIS) for voluntary reports by cockpit, cabin and ground personnel.

Many databases include some human performance data. Databases with emphasis on human performance include

- Aircrew Incident Reporting System (AIRS; confidential human factors reporting system for Airbus aircraft),
- Confidential Human Factors Incident Reporting Programme (CHIRP; independent and confidential reporting system for UK commercial and general aviation industry),
- Aviation Quality Database (AQD; integrated set of tools for safety management and quality assurance),
- Aviation Safety Reporting System (ASRS; FAA-NASA system for voluntary reports by pilots, controllers and others),
- Computerised Human Error Database for Hum Reliability Support (CORE-DATA; general database on human errors and incidents, which have at first primarily been collected in the nuclear industry, but now also includes other industries), and
- NLR Air Safety Database (large database containing a collection of aviation incident/accident databases as well as occurrence and flight operational data, including airport databases, flight exposure data, weather data, fleet data, and more)

Databases containing aviation safety data and information (e.g. accidents, incidents and occurrence data) can support safety assessments by identifying human performance issues and its variability and by providing data to estimate the frequency of occurrence for particular events. Moreover, linking potential causal factors to exposure data (e.g. number of movements, environmental conditions, operational environment) enables the analyst to estimate event probabilities and conditional event probabilities. To do so, the analyst specifies a database query for specific causal factors or events in accidents, incidents and occurrences, e.g. human or organizational factors (if available in the database). The analyst may use a data mining tool to support the data search. Data from a database query must be further analysed to determine the data quality, applicability, etc. In order to determine a frequency of occurrence the analyst must divide the number of applicable events by denominator data (exposure data) describing the number of cases in which this event might have occurred. The outcome can be expressed in

different formats, e.g. per flight, per flight hour, per landing, per operational environment (e.g. day/night, per type of approach).

Quality of the data is paramount in the quantification and analysis effort. The level of detail in the database records determines the level of detail of the human performance that can be analysed and quantified. In database records information on the event (what happened) is usually provided. Accidents and serious incidents are investigated by safety boards, implying that information on causal factors, human performance and organizational aspects are generally available. However, many minor incidents and occurrences are reported without further analysis into the root causes. In such cases it may be difficult to find out human performance or organizational aspects. As a result the level of detail at which certain human factors can be quantified may differ across different databases.

Limitations of using databases for identification and quantification of human performance or organizational aspects include the following:

- The results may have insufficient statistical significance. This may be the case if the database does not contain a sufficient number of recorded events or if the amount of exposure data is insufficient. In such cases, the results are often called a “best guess” and used as such in safety assessments.
- Data records within a database and across different databases may have different levels of detail. Depending on the level of detail and available information the analyst could identify human performance issues and determine the associated frequency of occurrence.
- Since databases store ‘reportable’ events, they present collections of recognised deviations from normal operations. Events that are not recognised or judged as safety relevant or otherwise not reported are not included in a database. Some events or operational conditions may not be covered by the database, e.g. the probability of icing conditions in flight, but may be available through flight operational data from airlines or other data sources.
- The condition in which the event occurred may not have been reported or may not be collected otherwise, so that data on the frequency of exposure to that condition could be unknown. For example, one may be able to determine the frequency of a certain flight crew error in a specific condition by means of a database; however, the frequency of occurrence of that particular condition may be unknown and could be very relevant for the analysis.

A.6 Human Factors Case

Eurocontrol has developed a Human Factors Case approach to provide a comprehensive and integrated approach for human factors aspects in order to ensure safe and efficient performance of an organization (Shorrock et al. 2004, Mellett and Nendick 2007). Human Factors considers aspects for designing technical and work systems, tasks, objects and places for people, within a socio-technical organizational context.

In the context of the current report, the Human Factors Case is especially relevant because it provides a link between studies on safety and human performance beyond the human error concept. Such interaction is especially useful in systemic accident models, which consider the variability of human performance in a broad context, rather than in a failure context only. Figure 21 illustrates this link by comparing potential issues that may be addressed in Safety Cases and Human Factors Cases. For instance, it illustrates the relevance of situation awareness, workload, team roles and training both from safety and human factors perspectives.

The Human Factors Case documentation (Mellett and Nendick 2007) provides guidance for human factors integration in five stages:

1. *Fact Finding*. Recording of factual information about a project, including its background, system and environment, key stakeholders and documentation. The objective is to scope the project from a Human Factors perspective to identify what will change, who will be affected, and how they will be affected.
2. *Issues Analysis*. Identification and prioritisation of the project-specific Human Factors issues and their potential impacts on the project. These issues are classified into six main categories:
 - Human in system,
 - Organization and staffing,
 - Procedures, roles and responsibilities,
 - Teams and communication,
 - Training and development,
 - Working environment.
3. *Action Plan*. Development of an action plan that describes actions and mitigation strategies to address the Human Factors issues identified for the project.
4. *Actions Implementation*. Implementation of the action plan, with as output the Human Factors Case report that describes findings and conclusions from the actions taken to address the Human Factors issues.
5. *Human Factors Case Review*. Independent assessment of the Human Factors Case.

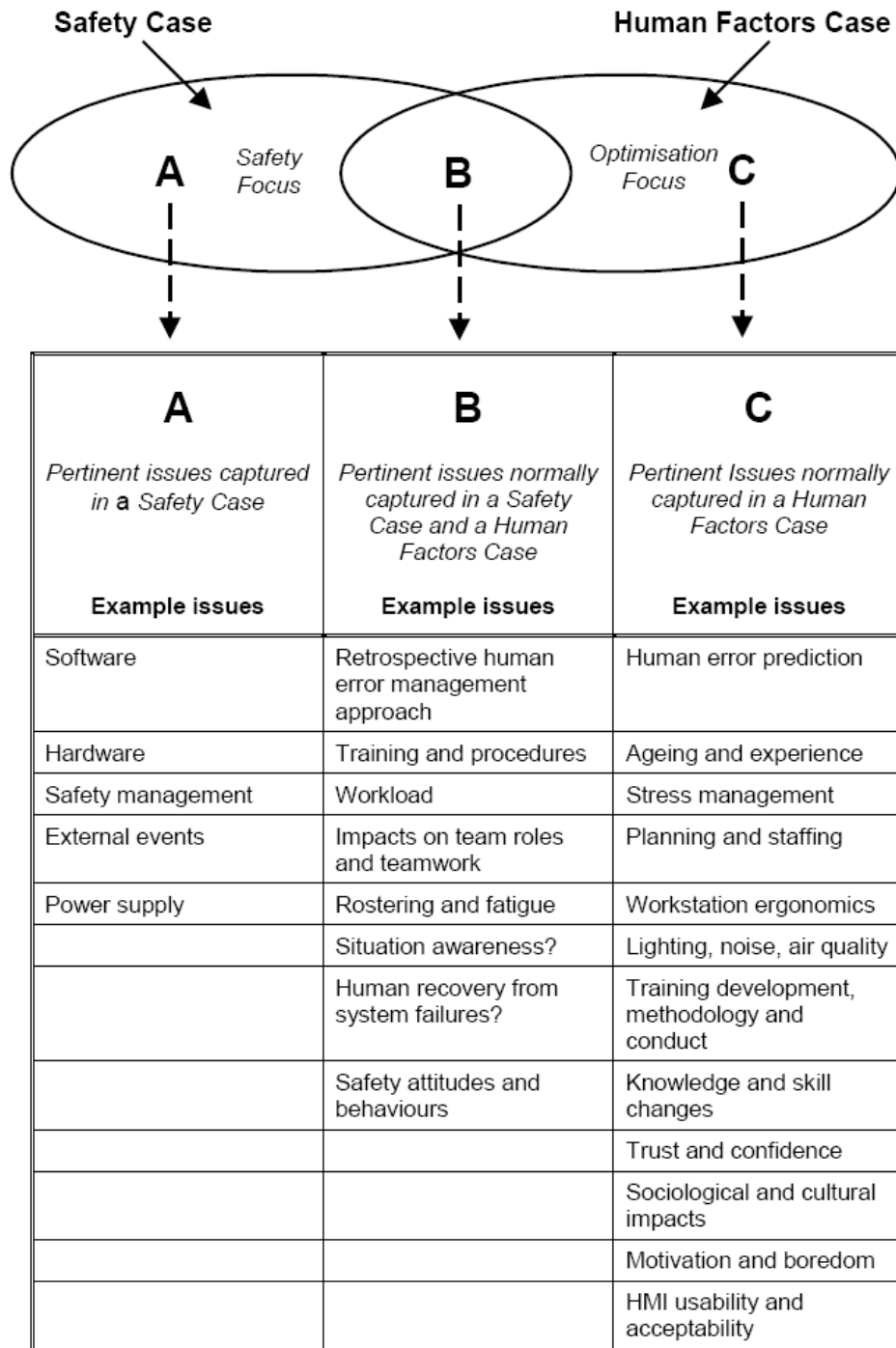


Figure 21: Comparison of issues typically captured in safety cases and human factor cases (Shorrock et al., 2004).

A.7 Tripod

Tripod is an accident model that describes the interactions between three components (Hudson et al., 1994; Groeneweg, 1998) (see Figure 22):

- General Failure Types (GFTs) which represent latent failures (they are also called Basic Risk Factors in (Groeneweg, 1998));
- Unsafe acts of human operators in combination with hazards (triggering events);
- Accidents / incidents.

A main aim of Tripod is to help understanding the effects of organizational deficiencies, which are represented by in a taxonomy of 11 General Failure Types (see Table 12).

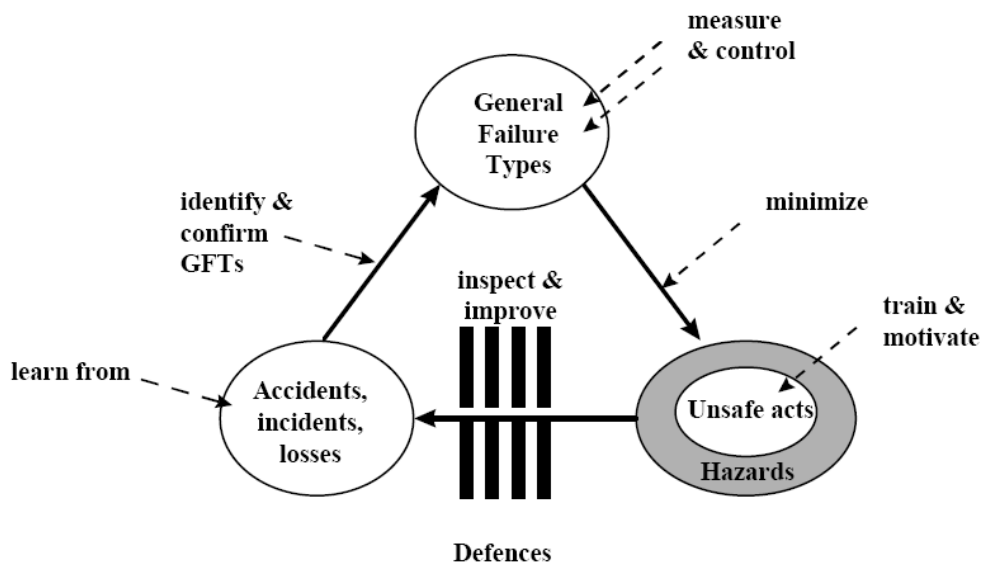


Figure 22: The three components of the Tripod accident model (Shell, 1997) .

Table 12: Description of Tripod General Failure Types (or Basic Risk Factors) (Groeneweg, 1998).

General Failure Type	Description
Design	Ergonomically poor design of tools or equipment.
Tools and equipment (hardware)	Poor quality, condition, suitability or availability of materials.
Maintenance management	No or inadequate performance of maintenance tasks and repairs
Housekeeping	No or insufficient attention given to keeping the work floor clean or tidied up.
Error enforcing conditions	Unsuitable physical conditions and other influences that have a disadvantageous effect on human functioning.
Procedures	Insufficient quality or availability of procedures, guidelines, instructions and manuals.
Training	No or insufficient competence or experience among employees.
Communication	No or ineffective communication between the various sites, departments or employees of a company or with the official bodies.
Incompatible goals	Friction between optimal working methods according to established rules, on the one hand, and the pursuit of production, financial, political, social or individual goals on the other.
Organization	Shortcoming in the organization’s structure, organization’s philosophy, organizational processes or management strategies, resulting in inadequate or ineffective management of the company.
Defences	No or insufficient protection of people, material and environment against the consequences of organizational disturbances.

The Tripod accident model forms the basis for the following two tools:

- Tripod-BETA is a retrospective tool for incident/accident investigation;
- Tripod-DELTA is a prospective tool for identifying weak spots in safety management.

Tripod-BETA

Tripod-BETA is a software tool that results in a tree-like overview of the conditions of the accident/incident investigated, with special emphasis of the failing defences and controls in terms of the General Failure Types. The tool prompts the user for information about the occurrence, which enables selection of the appropriate General Failure Types.

Tripod-DELTA

Tripod-DELTA is a survey-based approach with industry-specific questionnaires, which cover the 11 General Failure Types of the Tripod accident model. It results in indicators on the level of control/quality of organizational issues considered by the various GFT categories. These survey-based indicators can be presented in various ways, e.g. for groups of employees, for groups of companies, per time period, etc. It is not known whether there are dedicated Tripod-DELTA questionnaires for the aviation industry.

A.8 STAMP

It is recognised by Leveson (2004) that often applied sequential accident models, which explain accidents in terms of multiple events sequenced as a chain over time, and related reliability engineering techniques do not effectively account for (1) social and organizational factors in accidents, (2) system accident and software errors, (3) human error, and (4) adaptation over time. To account for these aspects, Leveson (2004) presents a model based on system and control theory: STAMP (Systems-Theoretic Accident Model and Processes). In the underlying concept of safety, accidents occur when external disturbances, component failures, or dysfunctional interactions among system components are not adequately handled by the control system. Here, the terms systems and control are used in a broad context, referring to all aspects and levels of a socio-technical organization. An accident is not understood in terms of a series of events, but rather as the result of a lack of constraints imposed on the system design and on operations. Safety is an emergent property from system components interactions within an environment. The emergent properties are controlled by a set of constraints (or control laws) on the behaviour of the system components, and accidents are the result from inappropriate constraints.

This safety concept has led to the continuing development of system modelling methods for risk analysis, which aims at understanding the dynamic risk contributions of technical and organizational factors, where ‘dynamic’ refers to the risk effects of changes in organizational

and technical factors over time. In such analysis, models of relevant organizational aspects are developed and the safety properties emerge from the interactions of these sub-models. For example, in a model for NASA’s safety culture sub-models such as Launch Rate, System Safety Resource Allocation, Perceived Success by Management, and System Safety Status were used (Leveson et al., 2005). In this example, emergent properties include perceived concern for performance, perceived concern for safety, level of risk, and fraction of corrective action to fix systemic problems. Another example of an application is the analysis of organizational processes in a water contamination accident (Leveson, 2004).

Archetypes for organizational safety based on control systems thinking are presented in (Marais et al., 2006). It presents (high-level) control diagrams for organizational issues, such as stagnant safety practices in the face of technological advances, decreasing safety consciousness, unintended side-effects of safety fixes, fixing symptoms rather than root causes, and vicious cycle of bureaucracy. As an example, the control diagram of vicious cycle of bureaucracy is shown in Figure 23. It consists of a reinforcing loop $R_{bureaucracy}$, which leads to a continuing increase in bureaucracy by management trying to improve (safety) performance by formalisation, a balancing loop $B_{control\ devices}$, which tries to regulate the performance by control devices, and a balancing loop $B_{make\ them\ feel\ better}$, which tries to counteract dysfunctional reactions in an organization by human relation treatments.

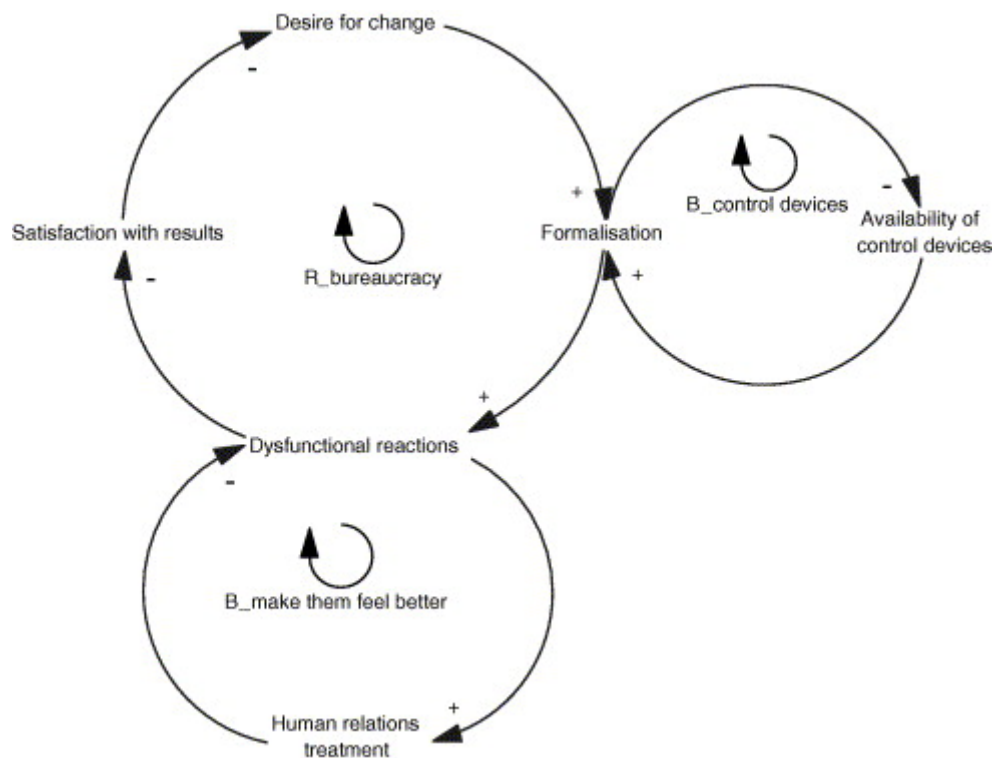


Figure 23: Control diagram for vicious cycle of bureaucracy (Marais et al., 2006)

A.9 TOPAZ

Motivated by stochastic system and control theory, researchers at NLR have developed a systemic accident model methodology for the evaluation of air traffic risk. This methodology uses Monte Carlo simulations and uncertainty evaluations to analyse the safety risk of air traffic operations. In (Blom et al., 2001a,b, 2003a,b) an initial version of this methodology has been introduced under the name TOPAZ (Traffic Organization and Perturbation AnalyZer). Subsequently, this methodology has been extended with multi-agent situation awareness modelling (Stroeve et al., 2003), an integrated qualitative safety risk assessment cycle (Blom et al., 2006a), risk bias and uncertainty assessment (Everdij et al., 2006a), and compositional specification of accident models by Petri nets (Everdij et al., 2006b).

An overview of the steps in a TOPAZ safety risk assessment cycle is given in Figure 24. In step 0, the objective of the assessment is determined, as well as the safety context, the scope and the level of detail of the assessment. Step 1 serves to obtain a complete overview of the operation. Next, hazards associated with the operation are identified (step 2), and aggregated into safety relevant scenarios (step 3). Using severity and frequency assessments (steps 4 and 5), the safety risk associated with each safety relevant scenario is classified (step 6). For each safety relevant scenario with a (possibly) unacceptable safety risk, the main sources contributing to the lack of safety (safety bottlenecks) are identified (step 7). The main results of the risk assessment cycle are the assessed risk levels and the identified safety bottlenecks. These results support decision making about the acceptability of the operation and identification of mitigating measures or improvements in the operation design. A more detailed discussion of the processes in these steps is provided in (Blom et al., 2006a).

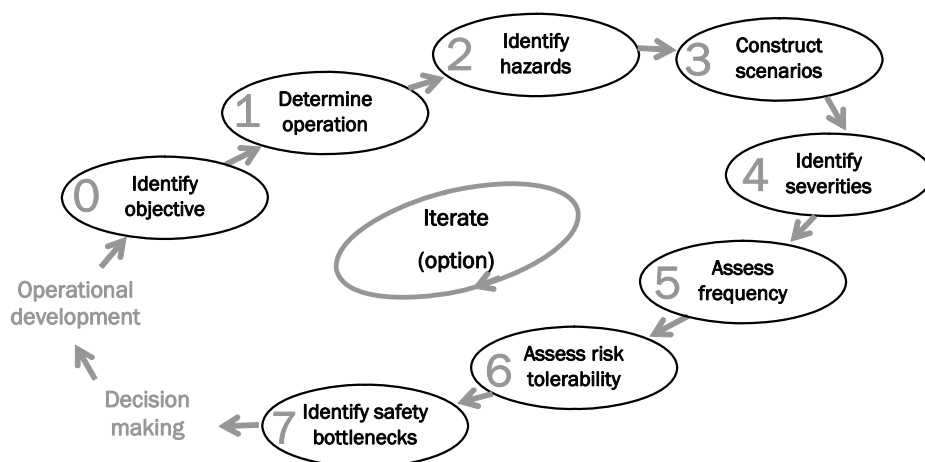


Figure 24: Steps in TOPAZ safety risk assessment cycle

In support step 5 of the safety risk assessment cycle, Monte Carlo simulations of accident models of air traffic scenarios are done. These accident models uniquely define the stochastic dynamics of the related agents (human operators and technical systems) by a compositional specification approach using a stochastic dynamic extension of the Petri net formalism (Everdij and Blom, 2006b). Within this Petri net formalism a hierarchically structured representation of the agents in the air traffic scenario is developed, including:

- Key aspects of the agents, e.g. situation awareness / task performance / task scheduling of a human operator, flight phases / performance modes of aircraft, or availability / status of an alert system;
- Modes within the key aspects of agents, e.g. task performance of a controller is monitoring / clearance specification / alert reaction, flight phase is taxiing / take-off run / rejected take-off / hold, or system availability is up / down;
- Dynamics within modes, e.g. the time needed for task performance, or the acceleration profile during take-off run, or the duration of an alert;
- Interactions between modes within key aspects, e.g. the transition to a next task, the transition to another flight phase, or a change in the availability of a system;
- Interactions between key aspects of an agent, e.g. the effect of situation awareness on task performance, the effect of an engine failure on a flight phase, or the effect of availability on the status of an alert;
- Interactions between agents, e.g. the effect of task performance of a pilot on the flight phase of an aircraft, or the effect of an alert on the situation awareness of a controller.

Here, the dynamics and interactions include deterministic and stochastic relationships, as is appropriate for the human performance or system considered.

In this systemic accident modelling approach, the performance of human operators and technical systems in an environment is thus represented in an integrated way by coupled stochastic dynamic models. The human performance modelling approach followed in TOPAZ is based on a contextual perspective in which human actions are the results of the interaction between human internal states, strategies and the environment (Hollnagel, 1998; Wickens and Hollands, 1999). The model for task performance of a human operator accounts for multiple tasks, which may be performed sequentially, based on a task hierarchy, or concurrently. The human performance model uses the contextual control modes concept from Hollnagel (1998), which distinguishes a number of control modes that impact the characteristics of performance of a human operator, e.g. the typical duration of a task, the choice of a task, or the probability that information is neglected. The control modes (e.g. opportunistic, tactical) describe a continuum from little to sophisticated control and mode changes depend on contextual or task-specific conditions. In this approach, human performance aspects are often represented as probability distributions, e.g. task duration, task accuracy, time until task performance, etc. In such

performance modelling, parameter values are based on operational observation, real-time simulation, expert interviews or related human performance models.

Another important aspect of TOPAZ is multi-agent situation awareness. The concept of situation awareness addresses perception of elements in the environment, their interpretation and the projection of the future status (Endsley, 1995). In an air traffic environment with multiple human operators, these aspects and associated errors of situation awareness depend on various human-human and human-machine interactions. The developed multi-agent situation awareness model describes the situation awareness of each agent (human operator, technical system) as time-dependent information of other agents, including identity, continuous state variables, mode variables and intent variables (Blom and Stroeve, 2004; Stroeve et al., 2003). Achieving, acquiring and maintaining situation awareness depends on processes as observation, communication and reasoning, which are part of the tasks of the human operator model.

As an integrated part of TOPAZ, a bias and uncertainty assessment method has been developed (Everdij et al., 2006a). This method supports identification of differences between the Monte Carlo simulation model and reality, and subsequent evaluation of the bias and uncertainty in the risk due to these differences. This evaluation includes assessment of the size of the differences and the associated risk sensitivity. Typically, feedback from operational experts is an important source of information in the bias and uncertainty assessment.

TOPAZ has been applied to a range of air traffic safety studies, some prominent examples are:

- collision risk analysis of parallel en-route lanes (Blom et al. 2003a,b,c),
- collision risk analysis of simultaneous missed approaches on converging runways (Blom et al. 2003d),
- collision risk analysis of incursion on active runway crossings (Stroeve et al. 2003, 2006)
- collision risk analysis of airborne separation assurance system-based (free flight) concepts (Blom et al. 2006b).