Using Complexity Theories to Reveal Emerged Patterns that Erode the Resilience of Complex Systems

Nikos Zarboutis[1] and Peter Wright[1]

[1] Human-Computer Interaction Group, Department of Computer Science, University of York, Helsington, York YO10 5DD, UK
nikosz@cs.york.ac.uk pcw@cs.york.ac.uk

Abstract. The objective of this paper is the demonstration of how complexity theories can be used to capture emerged patterns that appear during the evolution of a complex system and erode its resilience. After briefing on the primitives of Complexity Science and its use in the study of sociotechnical systems, some of these patterns are presented, describing both the mechanisms that favour their emergence as well as the ways that they can erode the resilience of a complex system. Finally, using an example from aviation, we illustrate how the previous concepts appear in practice, presenting evidence on the usefulness of this approach.

1 INTRODUCTION

Normal Accidents (Perrow, 1984) still happen! The Columbia accident was the latest disaster that could actually be said to be of this kind. The investigation report evoked the original work of Perrow (ibid.) to account for the characteristics of the mechanism that led to it (CAIB, 2003; p. 187; emphasis added):

"By their very nature, high-risk technologies are exceptionally difficult to manage. Complex and intricate, they consist of numerous interrelated parts. Standing alone, components may function adequately, and failure modes may be anticipated. Yet when components are integrated into a total system and work in concert, unanticipated interactions can occur that can lead to catastrophic outcomes".

Thus, despite the progress in technology, the methodological and theoretical advancements, it seems that complex systems persist on giving rise to large scale accidents. The nature of such accidents is different from typical events that can be attributed to individual failures of malfunctioning components, which can be adequately dealt with by typical decomposition techniques. Thus, while the reliability of individual parts has been significantly augmented, and often *components have been designed to "stand-alone" so as to be unaffected by problems of other elements of the system (e.g. a power failure in an aircraft*; Coombs, 1989), a new class of accidents has come to the forefront where large scale failures occur as the product of interactions among individually reliable parts. In other words, instead of talking about individual failures, we are now talking about *emergent, systemic failures*.

The classical socio-cognitive approaches to system reliability (i.e. human error studies; human reliability analysis; task analyses, etc,) have proven to be too limited in order to describe every aspect of a systemic failure, often resulting in misleading conclusions and remedial recommendations. Indeed, arguing with hindsight (Woods et al., 1994) what the human (or other) agents could have done gives us nothing but recommendations for patchwork (i.e. patches to close the "gaps" identified in a plan; Woods & Tinapple, 1999; Zarboutis & Marmaras, *in press*). As Woods et al. (1994, p.14.) put it "erroneous actions are symptoms, not causes" emphasising the view that humans adapt to a dynamic environment, instead of carrying out pre-specified plans. No matter how many patches there are, complex systems evolve in time and they keep on providing us with evidence of their amazing capacity to give rise to novel unanticipated situations.

Woods and colleagues (e.g. Cook et al., 1991), Rasmussen (1986) and others (see Woods et al., 1994; p. 18) have tried to raise the argument about the existence of invariants in sociotechnical systems and to look for recurrent patterns that lead to accidents. However, their value remained mainly descriptive, in terms of "regularities" in certain domains (e.g. healthcare) and they failed to capture the structural changes that link higher level patterns with systemic failures. Furthermore, such patterns could not be used for the enhancement of the system's adaptive capabilities, as they lacked a methodological framework that could link such structural changes to systemic failures, in order to come about solutions that could assure the elimination of the whole class of events that stem from them and to eventually turn them into design recommendations.

The objective of this paper is the demonstration of how complexity theories can help us capture such structural patterns which can emerge during the evolution of a complex system and weaken its resilience. Focusing on a few of them, that we have run into in many situations, we will provide evidence of how such patterns can give rise to emergent systemic failures, when there may not be an obvious individual failure. In order to do so, in this paper we will address the following questions:

- How can complexity theories and the relevant formalisms be used for the description of emergent phenomena in complex sociotechnical systems, where typical approaches often find it difficult to effectively address them.
- How and under which circumstances can co-adaptation emerge and what effect could it have on the resilience of the system as a whole?
- How can we go about the enhancement of the system's adaptive capabilities with respect to the control of emergence?

The structure of the paper is as follows. In the following section we present theoretical evidence on the use of complexity theories in the study of complex sociotechnical systems. Next we present some patterns that hamper the optimal adaptation of a complex system, defining co-adaptation and arguing on its effects to the resilience of such systems. In the following section, we proceed with the demonstration of this approach, using a specific example from aviation. Finally, we summarise on the approach, arguing on its usefulness in the pursuit of resilience in complex systems.

2 COMPLEXITY THEORIES & COMPLEX SOCIOTECHNICAL SYSTEMS

Complexity Science has received much attention in the past few years. B. Pavard and colleagues at GRIC/IRIT were rather the first to actually use Complexity Theories in the study of complex sociotechnical systems (Pavard & Salembier, 2003), aiming at both the understanding and the re-design of complex cooperative processes (ibid.), including studies in the SAMU emergency dispatch centre (Pavard et al., 1990), emergency rescue and training (Dugdale et al., 2000), air-traffic control (Salembier & Zuinar, 1998), etc. In the same line of research, Zarboutis & Marmaras (*in press*) used these notions building a method for the modelling of complex sociotechnical systems, based on the notion of Complex Adaptive Systems.

2.1 Evolution of a Complex System in Time and Space

Complexity Science describes the evolution of the complex system in terms of autonomous, interacting and co-evolving wholes. Each of them interacts with its neighbouring ones by triggering the other at its border. The latter, upon the reception of such stimuli, changes its internal organisation accordingly, through structural changes (i.e. adding or removing agents, interactions and combinations thereof) in order to assure the "survivor" of the individual parts (i.e. the state where each agent aims at the satisfaction of individual criteria, using local information and usually being unaware of the behaviour of the whole system), a process known as *adaptation* (Holland, 1992). This way, the stimuli presented to the interface of a whole do not determine the nature of the change in it; it is the internal dynamics of each of them that are responsible for the way that the whole and the agents that colonise it will self-organise, in order to adapt to these changes. Under the complexity paradigm, such internal processes can be described in terms of emergence, self-organisation and hierarchical control, as explained below.

Emergence & Hierarchical Control. Emergence is closely related to hierarchy (Checkland, 1984). The behaviour of the system is considered to be the product of local level interactions on various layers. On each of them, the components may locally interact; they can have their own structure and autonomous behaviour, which is not though a measure of the system's performance. On the contrary, there is always a higher level, where the properties of these interactions are evident and unique on that level. This operational mechanism of a complex system, where the product of local level interactions at a given level is evident at the higher one is called emergence and these higher level properties are called emergent properties of that level (ibid.; Simon, 1996).

This process of (bottom-up) emergence at the higher level though undergoes simultaneously a form of (top-down) hierarchical control that wants to assure that the product of emergence (i.e. the emergent properties) would be meaningful at the level of emergence. These levels of the hierarchy may represent various degrees of meaningfulness, in the form of either *order* (e.g. Prigogine & Stengers, 1993), or *meaning* (e.g. Winograd & Flores, 1986), *code* (e.g. Holland, 1992) etc., depending on the nature of the system and the processes involved. For two interlocutors for example, the product of their utterance can be considered as an emergent property at the level that gives to that discussion a unique meaning. This property (i.e. the verbal acts) has only meaning on the given level and an observer that fails to understand both the lower level and the level of emergence (e.g. focusing on the lower level as if the conversation is context-free) would be impossible to have an adequate understanding of the conversation, as to effectively engage in.

In that respect, out of the set of probable interactions at the base levels, only the possible ones qualify for emergence. The specific form of which of the possible will actually appear depends on the specifics of the system. In purely natural systems for example, Darwin's law of natural selection serves for such an explanation. In artificial, man-made, systems however, this is a great problem, as the more the options for emergence, the larger the uncertainty of a problem. The design of an effective hierarchical control system (e.g. a barrier system), so as to shape the emergent phenomena of the system, is the ultimate challenge for the engineering of complex systems. Accordingly, the resilience of a complex system is dependent on this form of optimal adaptation, which goes through the balance between emergence and hierarchical control.

3. PATTERNS THAT IMPAIR THE OPTIMAL ADAPTATION

While strong predictions (i.e. who, when, where, etc.) of the behaviour of complex systems are impossible to achieve, Zarboutis & Marmaras (*in press*) have pointed out that it is possible to identify some recurrent patterns, that once emerged, they actually divert the behaviour of a complex system towards a systemic collective event. Some of the most recurrent patterns mentioned above are self-reference, infinite loops, stigmergy, and co-adaptation as a combination thereof.

3.1 Self-Reference, Infinite Loops & Stigmergy

Self-Reference. Self-Reference is the pattern where an organisation (e.g. a set of agents) produces by itself, the structure that creates itself, in a recurrent way in time, as a system evolves. For example, in a previous study in the railways (Zarboutis & Marmaras, *in press*) self-referent phenomena were evident in the group of passengers, evacuating a flaming train. More specifically, the set of passengers, produce a crowd that (as a whole) poses influence on the way that the individuals behave in it and thus collectively forming (creating) it and this interplay between the individual and the collective drives the evolution of the crowd in time (along with other mechanisms of course; ibid.).

Infinite Loops. Infinite Loop, is a property of looping, i.e. the repetition of a set of processes until a condition is met. Once such condition is fulfilled then the loop exits to some given point. However, failing to meet this end condition, this iteration may continue endlessly and in most cases it can only be exited upon the imposition of some external force, intentional (e.g. the action of a human operator, if the system provides such affordance) or unintentional (e.g. a mechanical failure due to prolonged use of some piece of equipment). Infinite loops are typical errors resulting from complicated rule-books (i.e. the web of procedures used for the management of various events within an organisation), where a group of operators often find themselves in applying the same procedure again, after having completed a set of other ones.

Stigmergy. Two agents can interact either directly or indirectly. In the former way, the interacting agents exchange material, information or energy, in a straightforward fashion. According to Complexity Science though, an indirect mechanism of interaction is that of *stigmergy* (Grassé, 1959) where two parts interact, each modifying the environment, on which the other part adapts. Thus, under the co-evolving scheme described earlier, we could notice for example that for two parts belonging to different wholes (i.e. collection of agents), in a system and its environment, then the parts even if they do not have the ability to interact directly, they can do that through stigmergy. Such phenomena are evident in many types of complex systems, such as those in aviation that will be presented later in this paper.

These primitive patterns need not be mutually exclusive. On the contrary, our experience has revealed that they usually stem out of lower level interactions and collectively give rise to some higher level patterns, which shape the system's emergence diverting its operation towards systemic failures. Such a pattern that our experience has revealed in many cases is the so-called *co-adaptation*, which is responsible for some "peculiar" collective phenomena, not immediate apprehensible through analytical reasoning.

3.2 What is Co-Adaptation?

In general, *co-adaptation is the process where two agents (or agencies, depending on the chosen level of granularity) adapt to the same problem, each pursuing its own private goals*. For example, during a fire in a refinery, two agents can both adapt to the same problem (e.g. some leak), when someone chooses to cut a flow off from a pipeline locally, while some other controller may be trying to restore it centrally, each following a different plan to deal with the fire, each being unaware of the actions of the other. In the bigger picture, the agents need not be human; it could be a human interacting with a technological agent (e.g. a computer) where the human may pursue different objectives than the machine, for a number of reasons.

Within a complex organisation, co-adaptation is catalysed by stigmergy. However, when agents interact through stigmergy, co-adaptation need not have negative effects on system performance. Indeed, under normal circumstances, the typical outcome of a co-adaptive act could lead to *positive redundancy* (e.g. a double check). Nevertheless, when the interacting agents pursue different or conflicting objectives, co-adaptation can lead to unanticipated collective outcomes, given that the system as a whole provides relevant affordances for action. As a result, the system may collectively fail to adapt to external perturbations on the desired way and a systemic failure is likely to take place.

The elimination of such patterns requires the removal of the sources that lead to their emergence. Thus, if we can assure for example that the interacting agents would always form common objectives, or that the necessary external forces would be present as for an agent to exit the infinite loop that s/he is trapped into, then we will have achieved a more resilient organisation that would have the potential to create and maintain safety. Modelling the system as complex through an explicit account of emergence, self-organisation and hierarchical control, the role of such patterns can become immediately evident, while the relevant causes that erode the resilience of the system can be easily identified, as it will be shown through the presentation of an example in the following section of the paper.

4. AN EXAMPLE: THE NAGOYA ACCIDENT AS A SYSTEMIC FAILURE

On 26 April 1994 an Airbus Industrie A300B4-622R stalled and crashed during landing over Nagoya Airport, Japan. The technology employed in these *fly-by-wire* aircraft is such that the Flight Management Systems installed have the capacity of operating the control surfaces and the engines of the aircraft autonomously and apart from the will of the human pilots, via some automatic agents. In the case that we are going to present below, the resilience of the system, comprising such an intelligent aircraft, human pilots and a wider aviation system, was eroded by emergent patterns of interaction. In the remainder of this section we are going to present how the patterns defined earlier contributed to the erosion of the resilience of the system and ultimately in the emergence of such a systemic failure.

4.1 Factual Information

The critical events that led to the accident, as appeared in the investigation report, (Sogame & Ladkin, 1996) begin after the First Officer inadvertently triggered the *go-lever*. As a result the *Flight Director* switched from "land mode" to "go-around" mode, initiating a process to abandon the landing and gain altitude quickly and safely so that another attempts at landing could be made. The First Officer, although eventually aware that the lever had been triggered, tried to land the aircraft by applying "**nose down** input" and reducing thrust. However, with the go-around mode active, the more the pilot applied nose down force to the sidestick, moving the elevator accordingly, the more the autopilot moved the Trimmable Control Stabilizer in a **nose-up direction** to counteracts the pilot input because, as it was clear, it was still executing commands that would ensure a safe go-around (i.e. gain power and height). After a series of human adaptations and automatic adaptive responses, including the activation of the *alpha floor function* which increased thrust levels as well, the plane began to adopt a very steep climb. Despite the pilot's final attempt to actually go-around, when it was obvious that he could not land, the plane stalled and crashed over the runway!

The re-examination of these events under the prism of the complexity paradigm reveals that many **co-adaptive phenomena** took place**,** where **the human pilots** and the **autopilots** were trapped in a **self-referent situation** interacting **indirectly** through **stigmergy,** both **adapting on conflicting objectives.**

4.2 A Critical View on the Nagoya Accident under the Complexity Paradigm

Following Simon's (1996) criteria of *near-dissociability* in order to frame the system to be studied to ensure all relevant interactions are considered, the system for analysis is framed as in figure1a.

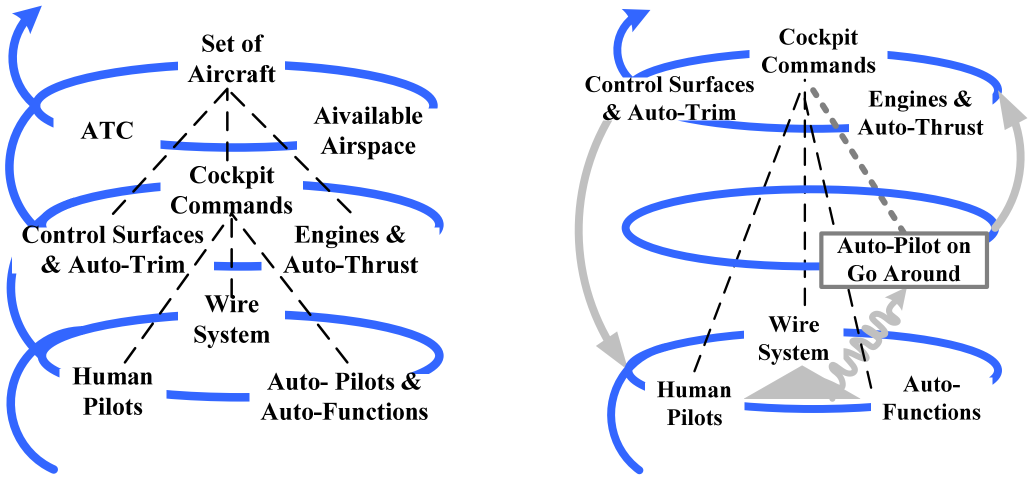

Fig. 1. (a) On the right: the normative model of the *Complex Adaptive System*, containing the following wholes: (a) the *Cockpit System* comprising the Human Pilots and the Autopilots System, (b) the *Control Surfaces System*, comprising the control surfaces (i.e. ailerons, elevators, trimmable horizontal stabilizers, slats/flaps), (c) the *Engines & Auto-Thrust System*, comprising the two engines and the automatic mechanisms that control these systems (d) the *Air Traffic Control System*, which in our case, due to the *practicality-validity agreement* can be reduced to Tower Control, (e) the *Available Airspace System*, delimited by the controllers and being occupied by the aircraft, (f) the *Set of Aircraft System*, reduced to those aircraft in the proximity of Tower Control, again due to the practicality-validity agreement. **(b) On the left: the model of the system as it evolved after the trigger of the "Go-Around" lever.**

During the *normal operation* of the system as a whole, at the bottom level the agents interact with each other, being triggered by changes across their border (e.g. a message by the EICAS). Within the *Cockpit System*, the interaction of the two pilots with each other and as a whole with the Autopilot is filtered by the *Fly-by-Wire System* (i.e. the imposition on the human pilots' inputs of the *flight envelope restrictions*). This process suffers a form of hierarchical control of the level above, so that the emergent properties would be unique and meaningful (i.e. that result to *airborness*) on that level and only. Hence, at the higher level of emergence, the interactions between the pilots and the autopilots result in certain *Cockpit Commands as emergent properties* that relate to the airborness of the aircraft. These commands arrive at the border of the Cockpit System as electric signals that interact with the two other systems (i.e. the Control Surfaces System and the Engines System), triggering changes at their borders. These signals thus, are imported in the other two systems and depending on the software logic embedded in the operation of the computers that control the servos that move the surfaces or the controllers of the engines turn these signals into actions. Thus, *on the level of airborness*, the triple co-evolution of (a) the *emerged cockpit commands*, (b) the *control surfaces system* and (c) the *engines systems* produces the synchronised positioning of the surfaces and the properties of the engines; in other words the attitude and the speed of the aircraft in the sky.

4.2 …damn it! How come like this? The Erosion of Resilience

Tracing the accident from the ground-up and using the model depicted in Figure 1a as a model of reference for the *world-to-analyse,* we can argue that the interaction between the human pilots and the autopilots, given the different objectives, resulted in such cockpit commands on a higher level which led to this systemic failure.

Stigmergy in the Cockpit. At some point during landing, the captain exclaimed: *"Damn it, How Come Like this?"*, expressing his frustration in front of the surprising effect that the aircraft as a whole was presenting him (i.e. a very steep pitch-slope, close to 53°!) as if **the human pilots were not making sense of the system**! Digging deeper into complexity we can see that the mode of interaction among the pilots and the autopilots had changed from "levelled and direct" to "non-levelled stigmergical" one, both acting and adapting to their common environment (i.e. the product of the triple co-evolution of the *Control Surfaces System,* the *Engines System* and the *Cockpits Commands* which was evident as the aircraft's speed and attitude), as shown in Figure 1b, above. The activation of the "go-around" mode led to an emerged –crystallised- structure, on an intermediate level between the levels that the pilots were standing and the plane as a whole (resembling Salthe's (1984, 1985, in Lemke, 2000) view of *downward causation* where an emerged structure may be relatively independent of the dynamics of the lower levels in a complex system). Although the system was designed so that the relation among the pilot and the auto-pilot could be levelled (as shown in Figure 1a), forcing out other options, the system did manage to exhibit intermediate behaviour!

Co-Adaptation in the Cockpit. Because of this intermediate behaviour, the emergent cockpit commands resulted because of the affordances that it provided its agents to pursue different objectives (i.e. the pilot to land the plane; the auto-pilot to go-around) and the means to adapt in order to meet them (i.e. to co-adapt) as shown in Figure 2.

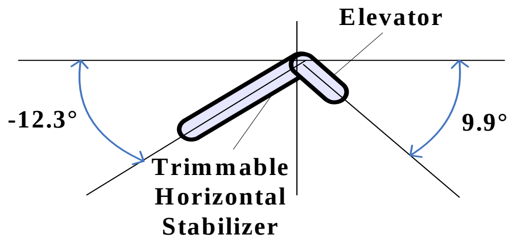

Fig. 2. The THS-Elevators relative positioning. The particular aircraft was built so that two control inputs for two different objectives could be allowed simultaneously in the pitch axis (Sogame & Ladkin, 1996), as: (a) The pilots could have the elevator moved through the side stick and the trimmable horizontal stabilisers using the trim wheels, (b) The autopilots could move the elevators and the trimmable horizontal stabiliser so as to act as Pilot Flying.

Under that co-adaptive scheme, the human pilot was changing the positioning of the elevators so as to bring the nose of the aircraft down, while the autopilot, reading the nose-down tense, was moving the THS on a nose-up direction, being entrapped in a **self-reference loop.** Overall, being unable to exit the loop, this process resulted in a situation where the pitch angle of the plane was determined by the aerodynamic effect that each surface had on the aircraft. Since the THS surface was approximately 3 times larger than that of the elevator, the emerged positioning shown in Figure 3 favoured the *nose-up* direction of the plane which eventually led to a stall and later to a crash.

5 CONCLUSIONS: THE PRACTICE OF RESILIENCE-ENGINEERING

In this paper we demonstrated how complexity science can examine some erroneous patterns that are known to divert the behaviour of a complex system to maladaptive collective performance and to a potential systemic failure. Co-adaptation and other relevant patterns were identified as factors that can impair the optimal adaptation of a complex system, eroding its resilience, as in the case of the Nagoya accident. In the bigger picture however, the concept of emerged patterns can have a direct repercussion in *the practice of resilience engineering* as they can be easily incorporated in HAZOP-alike techniques, that are still at the core of safety engineering. Typical inputs could be the following:

- How can two agents/ agencies/ wholes co-adapt?
- What is the probable effect on the system as a whole?
- How can we eliminate the emergence of such patterns?

The output of this process could point directly to the design of barriers, defences and other means that act on the emergence of unwanted properties. In this way, by identifying and removing patterns of suboptimal performance, we can go about the enhancement of resilience as we would remove the "sources of suboptimal adaptations" and not "the adaptations themselves" that may be useful in other times.

6. ACKNOWLEDGEMENTS

This research was funded by the EU FP6 RTN "ADVISES: Analysis, Design and Validation of Interactive Safety-Critical and Error Tolerant Systems".

REFERENCES

CAIB (2003). The Columbia Accident Investigation Board Report, Vol. I. Available as an electronic document at: *http://caib.nasa.gov/*.

Checkland, P. (1984). *Systems Thinking, Systems Practice*. Wiley.

Coombs, L.F.E. (1989). *Flight Decks and Cockpits*. In D. Middleton (ed.) *Avionic Systems*, Harlow, Essex, Longman, pp. 48-83.

Dugdale, J. , Pavard, B., Pallamin, N., El Jed, M. Emergency Fire Incident Training in a Virtual World. In Proceedings of the International workshop on Information Systems for Crisis Response and Management (ISCRAM 2004). 3-4 May 2004.

Grassé, P.P. (1959). La Reconstruction du Nid et les Coordinations Interindividuelles chez Bellicositermes Natalensis et Cubitermes. La théorie de la Stigmergie : Essai d'interprétation du Comportement des Termites Constructeurs. *Insects Sociaux* 6, 41-81.

Holland, J.H. (1992). *Adaptation in Natural and Artificial Systems. An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, 2nd edition. Cambridge, MA: The MIT Press.

Lemke, J. L. (2000). *Material Sign Processes and Emergent Ecosocial Organization*. In P. B. et al. (eds), *Downward Causation: Mind, Bodies and Matter*, Aarhus, Netherlands: Aarhus University Press, pp. 181-213.

Pavard, B. & Salembier, P., (2003). De l'ingénierie Cognitive à la Théorie des Systèmes Complexes. Un Parcours d'analyse et de Modélisation de l'activité Centrée sur la Conception. *Actes du 38ieme congrès de la SELF*, 24-26 Sept., Paris.

Pavard, B., Benchekroun, H. & Salembier, P. (1990). La Regulation Collective des Communications dans un Centre d'appel d'urgence: Analyse et Modélisation. In Actes du Congres IA, Biarritz, France.

Perrow, C. (1984). *Normal Accidents. Living with High-Risk Technologies*. Basic Books

Prigogine, I. & Stengers, I. (1984). *Order out of Chaos. Man's new Dialogue with Nature*. New York, NY: Bantam Books.

Rasmussen, J. (1986). *Information Processing and the Human-Machine Interaction*. North-Holland: Elsevier.

Salembier, P. & Zuinar, M. (1998). Analysing and Assessing Mutual Awareness in Cooperative Work Settings. ARAMIIHS-CNRS-UPS technical report, Toulouse, France.

Simon, H.A. (1996). *The Sciences of the Artificial*. Cambridge, MA: MIT press.

Sogame, H. & Ladkin, P. (1996). Aircraft Accident Investigation Report 96-5. Aircraft Accident Investigation Commission, Japan.

Winograd, T. & Flores, F. (1986). *Understanding Computers and Cognition, A New Foundation for Design*. Norwood: Ablex Publishing Corporation.

Woods, D.D. & Tinapple, D. (1999). W^3: Watching Human Factors, Watch People at Work. Presidential Address at *the 43rd Annual Meeting of the Human Factors and Ergonomics Society*, September 28, 1999, available at: *http://csel.eng.ohio-state.edu/productions/hf99.*

Woods, D.D., Johannesen, L.J., Cook, R.I. & Sarter, N.B. (1994).*Behind Human Error: Cognitive Systems, Computers & Hindsight*. CSERIAC SOAR 94-01 technical report, Wright-Patterson Air Force base, Ohio.

Zarboutis, N. & Marmaras, N. (*in press*). Design of Formative Evacuation Plans using Agent-Based Simulation. Accepted, to appear in Safety Science.