A STEP TOWARD A UNIVERSAL THEORY OF FAILURE HANDLING

Nadim OBEID*, Raj B. K. N. RAO**
*Department of Computer Information Systems, King Abdullah School for Information Technology,
The University of Jordan , JORDAN
**COMADEM International, 307 Tiverton Road, Selly Oak, Birmingham B29 6DA, UK

Summary

We explore, in this paper, some of the fundamental requirements needed for a Universal Theory of Failure Handling. We shall show that dealing with failure touches on our reasoning, predictive, evaluative and judgmental capabilities and thus it requires the ability to reason with incomplete and uncertain temporal information. It also requires reasoning with events before they even happen and about the effect of actions for as long as these are relevant, even if the available time does not permit. There may also be a need for reasoning about the reasoning process itself. We shall discuss the notion of failure with respect to decision-making and knowledge. We give a very brief presentation of Dorner's logic of failure and research into artificial intelligence and its implication for handling failures. We shall propose means of computing the degrees of failure induced by humans and in physical systems. In addition, we shall initiate a discussion on reasoning with failures and put forward a proposal for an integrative and proactive approach to monitoring, diagnosis and learning from failures.

Keywords: logic, failure, reasoning, decision-making, learning, diagnosis, monitoring, time, event, nonmonotonic.

1. INTRODUCTION

Failure seems to be a fact of life and almost every human endeavor (whether it is in business, politics, medical, defence, science, engineering, technology and management) has its way cemented by layers of failures (Rao (2006). It is everyone's ultimate goal to minimize the effect(s) of failure and to avoid, if at all possible. This phenomenon is shared by every discipline and/or every field of study and the patterns of behaviour to avoid or counter the effects of failure are quite similar. In addition to understanding a system, dealing with failure touches on our reasoning, predictive, evaluative and judgmental capabilities. It requires the ability to reason:

1. With *incomplete* and *uncertain* temporal information.
2. With events *before* they even happen.
3. About the *effect* of actions for as long as these are *relevant* and even if the available time does not permit.
4. About the *reasoning* process.

We aim, in this paper, to explore some of the requirements needed for a Universal Theory of Failure Handling. Section 2 is concerned with the notion of failure with respect to decision-making and knowledge. In section 3 we present a brief review of Dorner's logic of failure. Section 4 is concerned with research in artificial Intelligence and its implication for failure handling. In sections 5 and 6 we propose means of computing the degrees of failure by humans and in physical systems respectively. In section 7 we discuss the notion of reasoning failure. In section 8 we put forward a proposal for an integrative and proactive approach to monitoring, diagnosis and learning from failures.

2. FAILURE WITH RESPECT TO DECISION-MAKING AND KNOWLEDGE

A *decision* is a choice of an action from a set of alternative actions. Despite its long history and the interdisciplinary interest in this topical area, decision theory has nothing to say about the *nature* of actions and *how* they may become available. One of the key issues that must be properly tackled is *what* makes a good decision. They are formally characterized as actions that maximize *expected utility*. This view presumes that every action is associated with some outcome(s) and it is the responsibility of the decision maker to decide the nature of the outcome. However, there is some doubt about the validity of the utility-based approach regarding the possibility of estimating the outcome of actions (especially over *long periods* of time) (cf. Hare (1963)). There are many complex situations/scenarios that require the reasoner to have the ability to formulate decisions in *unforeseen* circumstances and to change the assumptions that underlie decisions. The adequacy of the required information is decided by the criteria of *preference* employed by the decision maker. Decisions that satisfy the conditions of ideal *rationality* may not have the *opportunity* to be made.

This is partly because, even with coherent preferences, it may take longer than permitted to obtain the relevant information or to decide what is reasonable. Thus, decisions may have to be made with *incomplete* information. Furthermore, information not vindicated by the preference criteria may not be considered even if it is relevant to the decision(s).

Practical reasoning is a *complicated* process. Human-based systems have theories or *perceptions* of the world, learn *new* information which may lead them to *update* their theories, make decisions using the *acquired* knowledge, have definite *goals* and *intentions*, and may act based on *some* of the decisions. It is important to note that each of these steps is *susceptible* to failure. The world-view may be false. Even if *beliefs* are true, decisions may be *inappropriate* and action may fail. Any of these `failure modes' can have a contributory role in a causal chain leading to a failure.

Failure occurs even in simple mundane tasks. The reasons for failure are numerous. Some of these reasons, just to mention a few, are:

1. Failure to *check* the appropriate information before acting.
2. Failure to check the *availability* of resources before a plan is carried out.
3. Making (inaccurate) *assumptions* and not *re-evaluating* their *validity* before acting.
4. Failure to consider other equally viable *alternatives.*
5. Failure to check for the correct order of a *sequence* of actions.

In complex situations and dealing with large complex systems, the tendency of *losing* sight/control is even greater. Some of the characteristics of decision-making in complex systems/situations are:

1. A failure in a complex system may not be *detectable* for long periods.
2. Complex systems/situations are *difficult* to evaluate.
3. A complex system can fail in numerous ways *(multicausal).*
4. The mode of failure of a complex system/situation is not easily *predictable*.
5. The crucial variables in a complex system/situation are not easily *identifiable*.
6. The more complex the system/situation, the greater is the possibility of *unexpected* failure.
7. Sometimes ideas are expressed *poorly* or are so *complicated* that those outside the group fail to *challenge* it because they may be seen as 'stupid' for not *understanding* it in the first place.
8. An *incorrect* assumption leads to incorrect *conclusion* or decision. In many situations, incomplete or misleading information may even become the norm. Assuming a piece of evidence just to force the conclusion to fit the facts is seldom correct. This only results in more accumulated failures or in 'passing the buck' to wrong individuals.
9. Most industrial accidents result either from human *behaviour*, such as bypassing safety devices or failing to ensure that a machine is in a safe state before entering a hazardous zone, or from 'systematic failures', such as incorrect selection of safety devices. In some situations humans can be the worst of the worst.
10. The more the world grows *'nano'*, the more the challenge becomes '*macro*'.
11. To understand the *root causes* of failure, it is important to comprehend the underlying *reasoning* and *decision-making* processes and why they sometimes fail miserably. The processes depend to a large degree on certain, correct and complete information and knowledge. However, *certainty* and *correctness* of information may need to be established and new knowledge may have to be discovered, generated and disseminated.

Reasoning and decision-making are closely associated with the gathering of the unbiased data and with the principles of accessing, manipulating and continuously evaluating intelligently different alternatives in order to choose the most appropriate for a given situation. Logic of different kinds is needed in order to correctly evaluate and manipulate information and knowledge. Logic is the ability, which allows us to generate knowledge from information. Development of *meta-diagnosis* through *meta-knowledge* will pave the way to successfully deal with failures.

3. THE LOGIC OF FAILURE

Dorner (1997) concluded that more often than not, humans completely fail to effectively manage complex systems. He states that:

1. Due to the slowness of our processing of conscious thought, we tend to make *shortcuts* in our decision-making (e.g, act before we have clearly defined what our goal is, or collected the required information).
2. We tend to *oversimplify* our models of complex systems by focusing only on one or two *key* variables and underestimating the importance of other interacting factors.
3. We are poor in *analyzing* and *forecasting* based on sequences of data in time. We tend to assume linear extrapolation of trends. We cannot cope well with accelerating or decelerating *changes*.

4. We tend to see new situations as simply *extensions of old*, established situations, and therefore apply old, established actions, which may not be appropriate.
5. We tend to *ignore* the possibility that actions we take now may have unintended side-effects, and may cause problems that currently do not exist.
6. We make *"ballistic"* decisions, where we do not monitor the outcomes of those decisions after we have made them.
7. We only act if we feel competent to do so. Without some expectation of success, we are likely to not act at all.
8. We form *simple hypotheses* and *limit* the search for information in order to preserve our own self-perception of competence.
9. We, sometimes, pursue planning, information gathering and structuring processes that go on interminably as a defense against the possibility that we are incompetent. This may keep us from making contact with the reality that our actions are not working.
10. We, sometimes and for self-protection, only solve those problems that we know we can solve, despite the fact that those may not be the most important or pressing problems.
11. We are not particularly effective at recalling past information and events, which can lead to us to repeat past and inappropriate decisions.

In more details, humans developed a tendency to deal with problems on an *ad hoc* basis (cf. P6): we fail because we tend to make a small mistake here, a small mistake there, and these mistakes add up (cf. p.7). He asserts that failure develops gradually according to its own logic. People court failure in predictable ways (cf. P.10) such as:

1. Acting without *prior analysis* of situations.
2. Failing to anticipate *side effects* and *long-term* repercussions.
3. Assuming the *absence* of immediately obvious negative effects means that the *correct* measures have been taken.
4. Being *blinded* to emerging needs and situational changes by *over-involvement.*

Dorner distinguishes between good decision-makers and poor decision-makers (cf p. 21). Good decision-makers have the following attributes:

1. They are more *sensitive* to subtle changes and act upon them more *aggressively*.
2. Their decisions take different aspects of the *entire* system into account, not just one aspect (cf. p.22).
3. They are more interested in the *causal* links behind events.
4. They often reflect on their own *behaviors*, make efforts to modify and structure them.

He cites the following as contributing to the faulty logic of decision-makers:

a) The tendency of a group of experts to reinforce one another's conviction that they are doing everything right (pp. 33-34).
b) We must often make do with tentative solutions because time-pressure forces us to act before we can gather complete information or outline a comprehensive plan. (p. 40).
c) Planners and decision makers may have no direct access, or indeed no access at all, to information about the situation they must address. (p. 40).

He defines a "system" as "a network of many variables in causal relationships to one another." (p. 73) He suggests that it is "wise when correcting a deficiency to consider it within the context of its system" and "considering the system ... means recognizing the different ways the variables can affect one another ". Among the categories into which he groups such interrelationships are *feedback* (positive and negative), *buffering*, and *critical versus indicator variables*.

He describes large bureaucratic organizations as well-buffered systems. The intricacies of their structures are well suited to negative feedback, and they are reflective of general goals that are widely shared. Thus, bureaucracies are well positioned to continue consuming resources even though they may be poorly structured to address the objectives underlying the generalized goals.

Regarding the incompleteness of information *needed to make a rational choice,* he seems to implicitly propose resorting to record events, as they happen in time, which can be communicated/revisited to reduce uncertainty. He also emphasizes the need to deal with changes over time with prediction of how they may extend into the future, taking into account additional contingencies. He warns us of excessive planning and information gathering as they keep us from making contact with reality where we will not know whether or not our measures are working.

Ultimately, he suggests, "There is only one thing that does in fact matter, and that is the development of our common sense." (p. 198) Yet, in contradiction to that assertion, he reiterates:
'Temporal configurations … often seem beyond common sense. … we do not give adequate attention to the characteristics of processes that unroll over time. (cf. p. 198).

4. ARTIFICIAL INTELLIGENCE (AI) RESEARCH AND FAILURE

Dorner seems to emphasize analysis and proper decision-making. The areas that closely capture the spirit of Dorner's suggestion in Artificial Intelligence (AI) are planning and reasoning about actions and changes. It is important to note that at least one area of research in AI, namely reasoning with incomplete information, is based on a notion that failure and progress was hampered by failure.

A classical representation of an AI planning problem is as follows: an agent, in an initial situation/state, has available a set of actions where an action can be viewed as a partial function that transforms a state into another and it is only operative on a state if its preconditions (a set of conditions) are satisfied. A planning problem then becomes a search for a series of realistic actions that successively transform the initial state, $S_{0,}$ into a goal state, S_G. Given a state S, the change is driven by the performance of actions such as, e, to result in a new state, $S_{1,}$ if, e, is operative on S.

There is a need to formalize *commonsense* reasoning. The best that AI workers could provide, so far, is formalizing micro-worlds that represent limited domains of knowledge and reasoning (cf. ([Davis (1991), Genesereth and Nilsson (1987)).

To check whether a plan has been successful, there is a need to check if the goal holds in the final state. This requires *predictive* reasoning: the ability to infer what may hold in later states that result from the performance of selected actions given the information about the initial/earlier states.

In addition to issues such as incomplete information, multiple agents, and continuous change, some of the problems associated with the formalization of action and change that are of interest to AI are: The Frame Problem, The Qualification Problem and the Ramification Problem.

4.1. The Frame Problem

The core of the frame problem is this: in any complex situation, there is a need to have to properly *frame* the problem in order to achieve a solution. This means identifying what is relevant and what is not, and determining the relevant information/knowledge needed to reach a solution. However, a proper framing of a problem requires a deep understanding of it.

For AI, the frame problem is the formalization of *inertial* reasoning; inferring what does not change when performing an action. Unlike many technically interesting problems that emerged in AI, it has attracted the interest of philosophers (cf. Pylyshyn (1987), Ford and Pylyshyn (1996)). It is open-ended, and can depend on a wide variety of circumstances. The purely logical Frame Problem can be solved using monotonic logic by simply writing explicit axioms (which may be infinite) stating what does not change when an action is performed. The non-monotonic solutions, which are based on **failure**, treat inertia as a default; changes are assumed to occur only if there is some special reason for them to occur. Absence of change is inferred when an action is performed unless a reason for the change can be found in the action axioms.

4.2. The Qualification Problem

It arises generally in connection with the formalization of commonsense generalizations. Typically, these involve exceptions, which may iterate endlessly. In a sense, this problem is addressed somehow by the non-monotonic logic systems by allowing commonsense generalizations to be formulated as defaults: the initial generalization can be stated as an axiom and qualifications can be added incrementally in the form of further axioms. It is important to note that not every non-monotonic logic provides adequate mechanisms for qualification. Default logic, for instance, does not give the intuitively desired conclusions.

Several aspects of the Qualification Problem remain challenging research problems in AI. For instance, no distinction is made between actions that cannot be attempted (probably because they are known to fail or unrealistic) and those that can be attempted, may fail. With the latter type of actions, those that may fail, we may need to reason about failure and its consequences. There is a need for a theory that can account for the ways in which actions (and plans that contain them) may fail.

4.3. The Ramification Problem

It is essentially to formalize the indirect consequences of actions, where "indirect" effects, which are not delayed but are temporally immediate and causally derivative. It is closely related to the Frame Problem (cf. Thielscher (1989, 1996, 2000) and Giunchiglia, Kartha and Lifschitz (1997)).

4.4. The Need for Explicit Causal Information

Hanks and McDermott (1987) showed that the existing systems of nonmonotonic logic (cf Reiter (1980), McCarthy (982)) were unable to give the right solution to the Yale shooting problem. Indeed, this example and other simpler examples (cf. Pearl (1988), Lifschitz (1990), Lin and Reiter (1994)) illustrated the need to include explicitly causal information in the input to defeasible reasoning.

Our knowledge of a situation is usally partial. Causal relations are no exceptions. It is, thus, difficult to conclusively confirm whether or not two arbitrarily selected facts are causally related. It seems that we usually apply a principle like Occam's razor: we assume that two facts are not causally related unless we have strong positive evidence that suggests otherwise. Thus, we may use a defeasible implication, that may fail sometimes, to capture causal relations.

5. HUMAN INDUCED FAILURES

Human beings seem to have the ability to solve complex problems. Such capability was more appreciated by the AI community when the attempt was, and still being made to capture such skills and expertise in automated intelligent systems.

The key for human success seems to lie in their innate ability to intelligently perceive, learn and reorganize their strategies and problem-solving heuristics. Saying that, we are admitting that failure is at the essence of such a process and it is unavoidably necessary to prevent failure in similar (complex) situations, where using analogical reasoning could be beneficial. However, given a novel complex situation and sufficient time pressure, it would be, according to the frame problem, highly unlikely for a single human being to identify the relevant features in time.

Such a statement is partly confirmed by Dorner's work. Looking at Dorner's logic without reference to, or justification as to why, humans behave as they do with regard to failure, we may conclude that some of the reasoning for failure are as follows:

1. Time pressure and process limitation may encourage us to act before we have clearly defined what the goal(s) is, or collected the required information.
2. Inappropriate modeling of complex systems.
3. Inability to identify new situations.
4. Underestimating the positive impact of an integrated monitoring and diagnostic system.
5. Underestimating the positive impact of recording past information, cases and events.
6. Underestimating the positive impact that learning could have on the reasoning process ability.
7. Inadequate use of a comprehensive theory of time and change to reason with.
8. Underestimating the frame, qualification and ramification problems.
9. Lack of culture and discipline leads to error-prone society.

Decision-making should possess adequate information together with the knowledge and experience required for evaluation and comparison. If not, it engages in search for appropriate information and required knowledge to carry out the process. However, extensive search will slow down the whole process to the point where it could be pointless. In other words, appropriately stored (and indexed) information and experience will facilitate the process of decision-making. The expectation is that the outcome of the process is the best conclusion that can be reached based on the available information. However, there may be difficulties and obstacles, which they somehow correspond to the reasons cited above, in reaching the best conclusion, such as:

a) Time pressure and/or the need to act before the required information, knowledge and experience are determined.
b) Incompleteness, incorrectness and/or uncertainty of information to be acted upon.
c) Limited reasoning ability to understand (complex) interactions between the different components of a system/situation and inability to make a proper manipulation and evaluation of information and/or the possible alternatives.
d) Lack of experience: not monitoring events and keeping records of previous encountered situations.This corresponds to (4) and (5) mentioned above.
e) Only react when conditions reach an alarming point: there is not much consideration of the health of the components and/or interactions between the different components unless abnormality is exhibited through the conditions being monitored.

Avoiding failure in complex situations boils down to addressing those problems in the following way:

Let DHF = Degree of Human Failure

IK = Degree of Insufficiency of Knowledge (e.g., incompleteness, uncertainty, inaccuracy).
TP = Time Pressure (e.g., in the sense that the issue has not been given adequate Consideration)
LRA = Degree of the Limitation of Reasoning Ability.
LE = Lack of Experience.
TBI = Tardiness before Interference.

$$HDF = IK*TP*LRA*LE*TBI$$

6. FAILURES IN PHYSICAL SYSTEMS

Humans play an essential role in systems failure, which seem to fail for reasons that are quite compatible with those of systems. However, compared with systems and given time, they always have an alternative solution or view on the problem whereas a system may have choices limited by the encoded knowledge.

Most operational systems are developed to realize a certain functionality. Thus, the approach can be roughly divided into the following phases: (1) Analysis where (functional and non-functional) requirements are determined; (2) Design whose outcome is an appropriate model that guarantees that the objectives of the system can be realized; and finally (3) the actual development where the model is materialized into the operational system. The approach is model-based; experiential knowledge regarding the system functionality and the interactions between the different components hardly comes into play.

Failure, in physical systems, is the responsibility of the systems developer/engineer whose role is to build systems, following an agreed specification that performs its intended functions throughout its operational life without fail. All systems, however, ultimately fail to perform its intended functions during its operational life at some stage. However, we may attempt to reduce the negative impact of failure and use it in a beneficial way as an opportunity for learning and understanding more about the system.

The primary causes of failures in physical devices can be ascribed to:

1. Incorrect assumptions, due to the incompleteness of knowledge, with regard to system requirements.
2. Lack of knowledge of/about the function of the device being operated versus its design. The different components that constitute the system may not properly be integrated.
3. Faulty design/structure. Different components that constitute the system are not properly integrated to achieve its functionality. Control may not be appropriate. Encoded knowledge is incomplete, incorrect and/or inappropriate.
4. Incorrect operations/user error: Poor estimation of the effect of some behavior or maneuvering of the device.
5. Lack of knowledge about appropriate maintenance time and procedure. This may result in material deterioration/failures.
6. Inappropriate conditions/environment and poor fit between systems and its environment.

Let DD = Design defect.
IK = Degree of Insufficiency of Knowledge encoded in the system (e.g., incompleteness, uncertainty, inaccuracy).
HF = Human Factor .
Ig = Degree of Ignorance (e.g., in the sense that the issue has not been under consideration).
W = Degree of Wear.
DOF = Degree of Failure.

$$DOF = DD*IK*HF*Ig*W.$$

7. REASONING WITH FAILURE

The previous two sections show, nearly in full agreement with Dorner, that the causes of most failures whether it be human or physical can be traced back to:

1. Insufficiency of available knowledge (e.g., incompleteness, uncertainty, inaccuracy) and its proper management.
2. Time Pressure (e.g., in the sense that the issue has not given adequate consideration).
3. Natural wear and tear (aging).
4. Construction/design defects.
5. Technological/Manufacturing defects.
6. Defects due to improper use (poor maintenance).
7. Defects due to variations in the usage.
8. Limitations in reasoning ability.

There is quite an abundant leterature that covers the above from various disciplines. However, little attention has been paid to the reasoning process and the possibility of its failure.

7.1. Towards Formalization of Practical Reasoning Processes

Practical reasoning processes involve many aspects such as determining reasoning goals, drawing inferences, making assumptions and evaluating alternatives. Furthermore, observations and/or tests that validate assumptions, decisions such as which reasoning goal to pursue and/or which assumptions to make, are basic component of a reasoning process. However, it is important to note that representation, knowledge and reasoning are entangled. It is, thus, not a straightforward matter to draw the line clearly between a failure and its associated cause(s) due to:

1. *Lack* of knowledge and/or misunderstanding.
2. *Misrepresentation* an/or inappropriate, characterization.
3. *Reasoning*.

It is quite reasonable to take the view that knowledge of a situation include:

1. Deep understanding and appropriate characterization: specifying its type, critical conditions that must be maintained etc.
2. *Manipulative* ability of desired actions together, with their preconditions and effects.
3. *Meta* reasoning and meta knowledge.
4. *Competencies*, which determine how well tasks are performed.

One way of formalizing the reasoning process would be to specify at each step:

- ➢ The rule that has been employed.
- ➢ Its justification (.g., the knowledge that has been employed).
- ➢ Its direct effects (e.g., the knowledge that has been created , what formulae, rules, and/or decisions have been enabled/disabled by the application of the rule, and so on).

A semantic account of the reasoning process can be given in terms of labeled transition systems. A transition system is a pair:

$$T = <S, E>.$$

Where:
S - is a non-empty set of reasoning states,
E - is a non-empty set of rules/events/actions (transitions) which act on S.

With every rule/event/action, $e \in E$, there is a transition, $\tau_{e,}$ such that $\tau_e(s) = t$, if the event/action, e, can transform the state, *s,* into the state, t.

In addition, a set S_0 of initial states can be specified.

Let AP be a fixed set of atomic propositions. Every state, $s \in S$, is assigned as a label the set of atomic propositions from AP that are true at s., i.e.,

$$L: S \rightarrow 2^{AP}$$

$$L(s) = \{A: A \text{ is true in } S\}.$$

The idea is that the atomic propositions (or what is hidden in them) describe adequately a reasoning state, including the values of all important variables and conditions.

Definition 7.1. A path (trace, history, execution), σ, in a transition system, T = <S, E>, is a sequence of states and events/actions which transform every state into its successor:

$$\sigma = s_0 \rightarrow_{e0}\ s_1 \rightarrow_{e1} \rightarrow_{e2} s_2\ ...$$

A path reflects a reasoning behavior regarding a goal.

7.2. Monitoring and Diagnosis of Reasoning Failure

It seems difficult to draw clearly the line between failures that are due to lack of knowledge, inappropriate use of knowledge and reasoning failure. In the same way that a system, which has the ability to learn has an advantage over a similar system that does not have such abilities. Also, a system that is not capable of monitoring and adapting its reasoning processes may be disadvantaged. Reasoning failures may mistakenly be considered knowledge failure. Furthermore, the system may not be able to predict and/or handle changes in its processes and/or environment and thus, incapable of adapting its knowledge and/or its reasoning to deal with the new circumstances.

In large and complex systems, there is a need to monitor the reasoning process and check its validity, and the validity of the consequences of decisions and to check their degree of conformity with what actually is expected. With the help of a diagnostic system and a meta-reasoning predictive ability of the reasoning process, we shall be able to determine whether a fault has occurred and how it could be corrected.

To be able to successfully carry out the process of detecting reasoning failure, there is a need for a description, in one form or another, of the expected reasoning behavior of the system. For instance, if the reasoning process is described in terms of rules, there will be a need for expectations, which may vary in their degree of generality and scope, about the effect of applying these rules such as: If this reasoning rules is applied, the value(s) which is (are) produced will satisfy so and so conditions. A fault would then result from a mismatch between the expected behavior and the actual behavior. However, the system needs to identify which expectations are currently relevant to the reasoning process.

It is important to make a clear distinction between expectation failures and reasoning failures. An expectation failure is usually caused by the occurrence of an event, which was not predicted by the system. It may be the result of a reasoning failure, if knowledge about the unexpected event occurrence was within the scope of the system's knowledge and the system could have planned for such an event.

Once the system has detected an expectation failure, it must be able to determine what point in the reasoning process the detected failure was created. This task is quite difficult, because there is no guarantee that the system will detect every reasoning failure at the instant it occurs.

8. MONITORING, DIAGNOSIS AND LEARNING FROM FAILURES

Most modern complex systems (such as AI systems) are increasingly knowledge-rich, dynamic and able to make decisions in complex and real-life environment. To be successful and to continue to be useful, these systems must be adaptable to the situations with which they have to deal, able to **meta-reason** about their reasoning processes, and to learn from their experiences. This is because, in addition to the complexity of the domains and the incompleteness of the knowledge encoded in these systems, it is not possible to predict the situations and many uncontrolled variables, which the system will have to deal with and the information required to deal with each of those situations.

It seems difficult drawing clearly the line between failures that are due to lack of knowledge, inappropriate use of knowledge and reasoning failure. In the same way that a system, which has the ability to learn has an advantage over a similar system that does not have such abilities. Also, a system that is not capable of monitoring and adapting its reasoning processes may be disadvantaged. Reasoning failures may mistakenly be considered knowledge failure. Furthermore, the system may not be able to predict and/or handle changes in its processes and/or environment and thus, incapable of adapting its knowledge and/or its reasoning to deal with the new circumstances.

Machine Learning (ML) systems (e.g., (Mitchell (1997)) are concerned with enhancing a system ability. However, they do not interfere with the reasoning processes, which are employed in manipulating a system's knowledge.

Obeid, Salah and Rao (2006) suggests that an enterprise/large system needs to properly manage its knowledge in order to achieve successful management of its assets, react appropriately to external demands, cope with the need for change and decide when to initiate change. They propose that to be effective, Condition Monitoring, Diagnostics and Assessment (CM-D-A) should be integral components of the Knowledge Management (KM) activity in an ongoing adaptive learning process.

In fact, it is not sufficient to detect, understand the reasons and how to recover from a failure. There is a need to employ meta-reasoning in order to learn from the failure and avoid future failures.

It has been argued in Obeid and Rao (2005) that there is a need for a finite-past temporal formalism to explain the cause of failure in a physical system, which may be due the occurrence of some unexpected event(s).

It has also been proposed (cf. Obeid and Rao (2002; 2004; 2005)) that reasoning with diagnostic temporal knowledge requires a formalism that:

1. *employs* an explicit representation of time and events;
2. *embodies* default rules; and
3. *incorporates* a domain description and the rules governing change to ensure that change to a state resulting from the successful occurrence of an event is minimal.

Monitoring and diagnostic knowledge is temporal, incomplete and uncertain. An Integrated Condition Monitoring and Diagnosis (ICMD) system has to work with a model of the system being monitored where the notion of a modeling itself suggests incompleteness and uncertainty. Temporal uncertainty could be exhibited in different forms and in many contexts.

In addition to time, events are associated with causes. An appropriate representation of events, have to go together with by a representation of changes in states, which can be captured by changes in the truth-values of fluents.

The role of an integrated condition monitoring and diagnostic system is exactly to witness that a property fails to hold and to decide on an appropriate action in order to detect the root causes of such unpredictable and chaotic behavior. Condition Monitoring (CM) can be equated with run-time verification, requires the ability to reason about past states. However, there is a need to reason about the future states and all possible paths in order to be able to predict/plan the system's behavior, which require knowledge of the temporal properties that must hold.

However, it is not sufficient to detect, understand the reasons and how to recover from a failure. There is a need to employ meta-reasoning in order to learn from the failure and avoid future failures. Improvement of reasoning processes can be achieved by continual monitoring and re-evaluation of their performance via the comparison between the actual and expected consequences. The system can adapt to new circumstances by noticing how the old ways of solving problems are inadequate and thus adjust its reasoning and its view of the world.

The laws of human reasoning should be closely related to the real situations as it is happening. The real systems are non-linear, non-stationary, multi-input and multi-output based, dynamic, fuzzy, chaotic and dynamic. As such, real situations are bound by imprecision and uncertainty and non-traditional methods of reasoning must be developed. Humans live in hope and reality. Like success failure is also a reality. As the complexity of human activity increase, uncertainties and poor-decision making confuses human mind, leading to chaos and failures of various degrees. A clear and intimate qualitative and quantitative relationship between human knowledge, set goal and intentions need to be established. This should tell us which course of positive actions to make in order to avoid failure at all costs. Humans are bestowed with innate meta reasoning and meta logic capability. Using this capability and by intelligently blending all the available tools, techniques and on-going developments (fuzzy logic, genetic algorithms, artificial neural networks, expert systems, pattern recognition, machine learning, hypertext, natural languages and high impact technologies), a robust and reliable solution could be found to minimise system's failures.

9. CONCLUSION

In 1973, Dr. Lofti Zadeh stated that "As the complexity of a system increases, our ability to make precise and yet significant statements about its behaviour diminishes until a threshold is reached beyond which precision and significance become almost mutually exclusive characteristics".

We have, in this paper, explored some of the fundamental requirements needed for a Universal Theory of Failure Handling. We have shown that dealing with failure touches on our reasoning, predictive, evaluative and judgmental capabilities and thus it requires the ability to reason:

1. with incomplete and uncertain temporal information;
2. with events before they even happen;
3. about the effect of actions for as long as these are relevant and even if the available time does not permit;
4. about the reasoning process itself.

We have discussed the notion of failure with respect to decision-making and knowledge. We have presented Dorner's logic of failure and research into artificial Intelligence and its implication for handling failures. We have proposed means of computing the degrees of failure induced by humans and in physical systems.. In addition, we have opened a healthy discussion on reasoning with failures and put forward a proposal for an integrative and proactive approach to monitoring, diagnosis and learning from failures.

10. REFERENCES

1. Davis, E., (1991), *Reasoning Common Sense*, San Francisco: Morgan Kaufmann.
2. Dorner D, (1997), *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*, HarperCollins Publishers.
3. Ford, K. M. and Pylyshyn, Z. (eds.), (1996), *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, Norwood, New Jersey: Ablex Publishing Co.
4. Genesereth, M. and Nilsson, J., (1987), *Logical Foundations of Artificial Intelligence*, San Mateo, California: Morgan Kaufmann.
5. Giunchiglia E., Kartha G. N. and Lifschitz V., (1997), Representing action: Indeterminacy and ramifications, *Artificial Intelligence*, Vol. 95, No. 2, 409-438.
6. Hare R. M., (1963), Freedom and Reason, Oxford University Press, Oxford.
7. Hanks S. and McDermott D., (1987), Nonmonotonic Logic and Temporal Projection, *Artificial Intelligence*, Vol. 33, 379-412.
8. Lifschitz V., (1990), Frames in the space of situations, *Artificial Intelligence*, Vol. 46, 365-376.
9. Lin, F., and Reiter, R., (1994), State constraints revisited, *Journal of Logic and Computation*, Vol. 4, 655-678.
10. McCarthy J., (1982), Circumscription - A Form of Non-Monotonic Reasoning, *Artificial Intelligence*, Vol. 13, 27-39.
11. Mitchell T., (1997), Machine Learning, McGraw Hill.
12. Obeid N. and Rao, B.K.N.., (2002), Innovative Trends in Knowledge Based Logical Reasoning in the Field of COMADEM, the Field, *International Journal of Condition Monitoring and Diagnostic Engineering Management (COMADEM),* Vol. 5, No. 3, 5-13, UK.
13. Obeid N. and Rao, B. K. N.., (2004), Diagnostic Temporal Reasoning in Model-Based Diagnosis (MBD) of Dynamic System, International Journal of Condition Monitoring and Diagnostic Engineering Management (COMADEM), Vol. 7, No. 1, 13-28, UK.
14. Obeid N. and Rao, B. K. N. (2005), Temporal Aspects in Condition Monitoring & Root Cause Failure Diagnosis of Modern Complex Systems, International Journal of Condition Monitoring and Diagnostic Engineering Management (COMADEM), Vol. 8, No. 3, UK.
15. Obeid N., Salah I. and Rao, B. K. N. (2006), The Role of Knowledge Management in Diagnosing & Prognosing System's Failures. Diagnostyka, No.1 (37), 9 – 16, Poland.
16. Pearl J., (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, Calif.: Morgan Kaufmann.
17. Pylyshyn, Zenon (ed.), (1987), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Norwood, New Jersey: Ablex Publishing Co.
18. Rao, B.K.N. (2006), Toward the Universal Theory of Failure, International Proceedngs of Condition Monitoring and Diagnostic Engineering Management (COMADEM), Published by Lulea University of Technology, Sweden. Pp. 85 – 101.
19. Thielscher M., (1989), Ramification and causality, *Artificial Intelligence*, Vol. 89, No. 1-2, 317-364.
20. Thielscher M., (1996), Causality and the qualification problem, in *KR'96: Principles of Knowledge Representation and Reasoning*, Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, eds., San Francisco, California: Morgan Kaufmann, 51-62.
21. Thielscher M., (2000), Representing the knowledge of a robot, in *KR2000: Principles of Knowledge Representation and Reasoning*, Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, eds., San Francisco: Morgan Kaufmann, 109-120.
22. Reiter, R., (1980), A logic for default reasoning, *Artificial Intelligence,* Vol. 13, 81-137.
23. Zadeh, L.A. (1973). The Concept of a Linguistic Variable and its Application to Approximate Reasoning, Memorandum ERL-M411, Berkeley, October.

Professor **Nadim OBEID** holds a B.Sc in Mathematics (Lebanese University, 1979) and a B.Sc. in Business Administration (Lebanese University, 1980). He also holds A Postgraduate Diploma (Essex University, 1982), M.Sc. in Computer Studies (Essex University, 1983) and a Ph.D in Computer Science (Essex University, 1987). In 1986, he joined the EUROTRA project as a Senior Research Officer and then in 1987, he took the post of a Lecturer in the department of Computer Science at Essex University. He took a leave of absence (without salary) from Essex University, in 1996, and became an associate Professor at Princess Sumaya University for Technology in Jordan where he was promoted in 2002 to a professor. In 2004, he joined, as a professor, King Abdullah II School for

Information Technology at the University of Jordan. The areas of research in which he is currently active are: Knowledge Representation, Multi-Agent Systems, Formalisation of Condition Monitoring and Diagnostic Reasoning, Knowledge Management and Logic of Universal Failure. He has been in the program committee of many international conferences. He has published around 20 publications in refereed international journals and more than 30 papers in edited books and refereed international conferences. He is the author of Winner of the Best Paper of the Year Award in the international journal of COMADEM (2001-2002).

Professor Dr. **Raj B. K. N. RAO** PhD, DTech is internationally recognised and respected for pioneering the development of the holistic discipline of Condition Monitoring and Diagnostic Management (COMADEM). His distinguished academic career spans nearly 40 years of teaching, research and consultancy expertise in the fields of Environmental Engineering, Dynamics and Control and Human Factors in Engineering. He is the Director of COMADEM International and since 1988 he has successfully organised a series of COMADEM International events. He has published a number of refereed technical papers / books / journals / conference proceedings and has supervised and examined numerous doctoral research programmes. He is currently Visiting Professor at a number of Universities including the University of Glamorgan and Vellore Institute of Technology, India.