USCW OCCASIONAL PAPER SERIES NO. 1 • AUGUST 2025

QUANTITATIVE RISK ANALYSIS

A Number-Free Introduction to the Method, with Examples Including Decision Support from Artificial Intelligence

By Elisabeth Paté-Cornell

EDITORS: Elizabeth Economy, Glenn Tiffert, and Frances Hisgen

US, China, and the World

A PUBLICATION OF THE HOOVER INSTITUTION

US, CHINA, AND THE WORLD

Quantitative Risk Analysis

A Number-Free Introduction to the Method, with Examples Including Decision Support from Artificial Intelligence

Elisabeth Paté-Cornell

INTRODUCTION

Risk can be defined as uncertainties about occurrences of undesirable events and their consequences. Risk analysis can be either quantitative or qualitative. The result of the qualitative approach is a judgment of how serious the risk can be. The result of the quantitative method is a failure probability, or the chances of potential losses. It is generally an input to a risk management decision. A risk analysis should be based on facts and should be independent of the preferences of the decision maker. Rational risk management decisions, by contrast, need to include both an assessment of the risk, and the decision maker's risk attitude, which may vary among people and organizations. The qualitative risk analysis method often mixes facts and preferences and reflects in words some knowledge but also the feelings of decision makers and experts. This common qualitative approach presents two major problems. It does not allow for a rational comparison of risks, and perhaps more importantly, it does not treat systematically the dependencies and trade-offs among different risk factors. The alternative is a quantitative probabilistic analysis, based on a complete set of known failure scenarios (or rather classes of scenarios), their probabilities, and the numerical values of their outcomes. The quantitative approach is the basis of this report. While avoiding equations and quantification, it is shown here how the quantitative method has been developed and applied in different types of settings. These could be sociological situations and processes, or problems of engineering system reliability including cyber risks. Three risk analysis examples are presented: the risk of anesthesia to a surgery patient, the risk of losing a space mission due to a failure of the spacecraft heat shield, and an AI-supported system of warnings and management of the risk of cyberattacks. An issue, when using an AI algorithm in risk management decisions, is the alignment of preferences. For the algorithm to be relevant, its risk attitude must match that of the actual decision maker. This implies that the AI system must be transparent enough about its risk preferences, and flexible enough to permit alignment of the risk attitudes between the AI system and the decision maker.

A Hoover Institution Essay

USCW Occasional Paper Series No. 1 • August 2025

RISK ANALYSIS

RISK: UNCERTAIN SCENARIOS WITH THE POSSIBILITY OF NEGATIVE OUTCOMES

Risk can be defined in several ways—for instance, qualitatively or quantitatively. It can simply be the probability of failure of a system or procedure in a given time frame (Paté-Cornell 2023).[1] Risk analysis is a quantitative method with numerical results. A qualitative risk assessment, however, generally yields risk results that can be simply presented as adjectives (the risk is "small," or "bigger than [another] risk").[2]

In a quantitative mode, the risk characterization includes computation of both the probability and the consequences of the failure scenarios (Kaplan and Garrick 1981; Paté-Cornell 2009). The risk can then be described by the probability distribution of the outcomes, i.e., by the chances that the losses might exceed different levels.

The risk analysis results are generally input to a decision under uncertainty. However, the preferences of the decision maker should not influence the risk estimates where they would introduce psychological biases. Instead, they should be part of the risk management phase, along with the risk analysis results (Paté-Cornell 2007b).

A simple but insufficient definition of a risk is often presented as the expected value of the outcomes—i.e., the product of the probability and the consequences of a hazardous event. While this result may provide useful information to a decision maker who focuses on the average losses, it is insufficient to support the decisions, for example, of a risk-averse decision maker, who puts more weight on the extremes than the expected value (Abbas and Howard 2015; Bier and Lin 2013).

WHY DO A RISK ANALYSIS?

Decision Support

Risk analysis provides a support to decisions under uncertainties in cases that involve some losses compared to the expected situation (Coombs and Pruit 1960). The ultimate goal is to protect systems, operations, and people, generally within cost constraints. One does a risk analysis to manage the risk-benefit trade-offs and optimize the use of risk management resources by anticipating potential problems and identifying the weaknesses of an operation or a system before an accident or a failure (Paté-Cornell 2022). In a seismic area, the analysis of the risk involves the frequency and severity of earthquakes, and its management for a homeowner relies on the strength of the house structure. Improving that structure to a given level of seismic security will involve a cost, which will be the owner's choice, and will depend on his or her financial means and risk attitude.

One key point of this report is the importance of recognizing and characterizing uncertainties, and quantifying failure probabilities in order to take timely and cost-effective risk

management measures. This requires recognizing and assessing not only the marginal uncertainties of relevant factors, but also the dependencies among events that can lead to a failure (Aven 2008).

The ultimate objective is to identify and support rational proactive risk management, accounting for all available relevant information, including precursors and near misses (Paté-Cornell 2004). These important data address protection not only against a particular failure mode, but also against other risks that may involve some of the same events and parameters. An example of ignoring prior information is that of the British Petroleum accident in the Gulf of Mexico in 2010, where several near misses that preceded the disaster were ignored. It was a difficult situation and a difficult well, but the crew hoped that since there had not yet been a catastrophic failure of the well and the blowout preventer, it would not happen on their watch. It did (US BP Commission 2011).

Rationality

Rationality is a key feature of the approach to risk management described here, and of the role of risk analysis as an input to these decisions. Rationality can be defined in several ways. A classic definition, which is used here, is based on a set of axioms designed by von Neumann about rational preferences—for instance, no circularity of choices (Abbas and Howard 2015). These axioms lead to the definition and encoding of a utility function for each of the possible outcomes, as assessed by the rational decision maker. In turn, this implies the choice of the option that maximizes his or her expected utility at any given time. That decision analysis approach ensures the consistency of preferences and the coherence of the decisions. The risk attitude is included in that utility function as a measure of how the preferences vary with the value of the outcomes (Lichtenstein and Slovic 2006).

TWO MAIN APPROACHES TO RISK ANALYSIS: QUALITATIVE AND QUANTITATIVE

Both the qualitative and the quantitative models are based on the identification of failure scenarios, but they do not process the information and characterize their results in the same way (Ostendorff and Paté-Cornell 2023).

Qualitative analysis is based mostly on expert opinions (Cooke and Shrader-Frechette 1991; Hora 2007). In the best cases, they are true experts who have the benefit of fundamental knowledge and can address situations for which there is no or little experience. Risk estimates can also be an expression of belief from people who may not have the same level of knowledge and/or whose heuristics involve biases (Kahneman et al. 1982). Some of them may be decision makers or advisors whose opinion is needed, sometimes immediately. Expert opinions reflect whatever they know about the problem but also, in many cases, their feelings, fears, and wishes (Lerner and Keltner 2001). Most risk management decisions in everyday life are made on that qualitative basis, and do not rely on more sophisticated considerations.

Because it is simpler, qualitative risk analysis is clearly popular, even in complex situations that might require a deeper understanding of uncertainties. It does not involve formal probability or a quantitative link to available evidence. The most likely scenario is often considered sufficient. That level of assessment is fine in situations where risk management seems obvious, and it is often unavoidable in an emergency.

If it is clear that a failure is going to happen, a decision has to be made on the spot, even though the situation could often have been considered ahead of time. Qualitative but informed decisions can then be made based on the opinions of experts when they are available. This may be the case in operating rooms where surgeons have gathered, through training and experience, the knowledge and information that allows them to face different kinds of problems, even new ones. The same is true of pilots in a difficult situation.

THE QUALITATIVE APPROACH: PROBLEMATIC ISSUES OF RISK COMPARISON AND EVENT DEPENDENCIES

When considering complex systems and situations, qualitative risk estimates often introduce some problematic issues that are seldom recognized. First, one cannot compare risks rationally and explicitly to make consistent and optimal decisions. A qualitative decision support is more likely to be biased than a well-defined numerical comparison. This may happen when a limited amount of resources must be optimally allocated to the management of different failure modes in a situation or a system.

More importantly perhaps, a qualitative risk analysis generally does not account for probabilistic dependencies. Some important situations may not be envisioned, and the chances of failure may be much greater than implied by an assumption of independence of the various failure modes. A typical case is that of "perfect storms," such as that which occurred in New England in October 2001 (Paté-Cornell 2012). Three storms converged off the coast of Maine, making their effect much more destructive than if they had occurred sequentially. As recommended by the US Coast Guard, most boats stayed ashore, but one decided to go to sea anyway and sank, drowning its crew.

AN ALTERNATIVE: A QUANTITATIVE PROBABILISTIC RISK ANALYSIS (PRA)

The basis of quantitative risk analysis is the explicit processing of uncertainties using probabilities and assessing the outcomes of the different scenarios independently from their probabilities (Garrick 2008; Paté-Cornell 2009). When a decision is made, both probabilities and outcomes for the different options are combined in a rational decision analysis framework (Abbas and Howard 2015).

The Method and the Factors of Quantitative Risk Analysis

Failure scenarios The first task is to generate a set of failure scenarios as complete as knowledge of the problem allows, in order to construct a legitimate probability distribution of the outcomes.

In the identification of failure cases, one generally needs to focus on classes of scenarios, rather than simple ones, since adding details to each would make their number uncontrollable and may not add useful information. The considered scenarios are the possible conjunctions of events leading to system failures, including technical failures of the system's components, human actions, and external events that may affect the whole system. Furthermore, the analysis focuses on the considered system at a given time, and one may need to anticipate its evolution in the future.

Bayesian probabilities Once the scenarios have been identified, their probabilities are computed per time unit or per operation. Short of relevant statistics—which is often the case—one uses Bayesian probability (de Finetti 1974; Apostolakis 1990), based on all information available for each component of the risk, including the judgment of experts.

A critical aspect of the probabilistic process is the treatment of dependencies among these factors. Consider, for instance, redundant subsystems. One of them (at least) has to work for their function to be performed. Their failures ought to be as independent as possible, and a risk analysis should include their correlation, if any.

Another key element of the failure probability is the effect of external events (for instance, floods) that affect all components of a system at the same time. Therefore, they create dependencies that need to be considered explicitly. An external event can be an earthquake shaking a building and all its components, thus increasing the structural failure probability (Cornell 1968). Similarly, in the political world, a coup d'état may affect simultaneously—differently, but with correlations—the different parties trying to capture power and may increase the risk of success of an insurrection.

The numerical part of the analysis involves the assessment of both the probabilities and the consequences of the different scenarios. That assessment leads to a probability distribution of the outcomes (generally losses), which will be a result of the risk analysis.

The extremes of the distribution must be represented in the computations, even if their probabilities look small. People who have sometimes ignored the possibility of a rare failure scenario before the fact have sometimes justified it after a disaster by arguing that it was not likely enough to be considered. Yet, the losses might have been controlled if the right measures had been taken in time. The decision to ignore a rare but grave event should be a serious one, and at least ought to be made explicitly. As part of risk analysis, resilience management is critical since it involves anticipating different loss scenarios and planning to avoid them even if they are the result of rare events. But mostly, the question is how to respond to them if a failure occurs and one is caught in a loss situation.

Loads and Capacities

Systems fail when the demand on them exceeds their capacity—i.e., their ability to absorb the load (Paté-Cornell 2009). Quantitative risk results are critical in the design and operation of a technical system, as well as in the planning and execution of a social, political, or commercial operation. In many cases, the failure of a scenario may result from excessive loads.

Therefore, to guide the design and management of a physical system, one needs to consider in a risk analysis the chances that in its operating life, the loads will exceed the capacity. For example, an analysis can show the probability that an earthquake load on a structure exceeds the maximum that it can take in a given time frame. One can test the structure's model on a seismic table, where the load is controlled as part of the test and the model represents the capacity of the real structure as well as can be represented given the scale. In the same way, one can assess the risk that a debt on a company exceeds its assets. In both cases, the result depends on all considered scenarios involving the loads and their probabilities at the time of the decision. An important part of the information about what can hit the system (the load) and what it can take (the capacity) includes the near misses and the partial failures that may have happened in the past, which provide information about the capacity of the system.

The Values and the Power of Bayesian Probabilities

Bayesian probability As mentioned earlier, probabilities of scenarios do not have to be based on statistics and often cannot. Bayesian probability (de Finetti 1974; Apostolakis 1990) is then the right tool. The events may have never occurred, in which case one has to consider the probabilities of conjunctions leading to possible failures. But past events may not represent future ones. For instance, the risk of a revolution in a country, stable so far, may not reflect its history. The structure of the scenarios and their probabilities are then based on knowledge about the factors involved, from past experience or expert opinions.

These judgments are thus the results of compounding relevant factors. To make such a judgment, one might use the fundamental Bayesian rule of logic: the probability of the conjunction of two events is the product of the probability of one, multiplied by the probability of the other given the first [$p(A \text{ and } B) = p(A) \times p(B \text{ given } A)$]. That simple logical rule is essential to all computations of the chances of scenarios and ensures that dependencies are properly accounted for.

Note that there are two kinds of uncertainties: aleatory and epistemic. Aleatory uncertainties represent the randomness in a situation where the probability of each event is well known but the outcome of each trial is not. For example, a die may be well balanced with a probability of 1/6 for each face but throwing the die will lead to one of the six possibilities with an aleatory probability of 1/6. Epistemic uncertainties reflect uncertainty about fundamental probability. In this case, it could be that the die is not balanced, and the probability of each face is not known. Statistics address aleatory uncertainties, but generally not epistemic ones. Bayesian probabilities are more general and allow assessing both types of uncertainties, providing a global probabilistic measure.

Bayesian Probabilities Rely on Multiple Data Sources

The choice of data sources is critical as it determines the quality and the credibility of the risk results. Not only should these sources be revealed and discussed, but all information that is needed should be acquired if time and resources permit.

The key data sources are as follows.

- *Actual, in situ data and statistics* For instance, flight data may be essential in the analysis of the risks involved in civil aviation. The data, in that case as in many situations, involve not only the performance of the different subsystems such as the engines, but also the external events such as storms that may affect the whole aircraft. One problem of external events is that they may introduce dependencies among subsystem failures. These events must therefore be considered in the design of the system or the process, and redundancies may be included to reduce the failure risk. Furthermore, the failures of redundancies must be as independent as possible for their combination to be effective, so that at least one of them can perform the task.

- *Surrogate data about similar subsystems elsewhere* Surrogate data may be a good source of information when actual data about the considered system are limited or inexistent. This may be the case for new ones, for which parts (similar but not identical) may exist elsewhere and may provide relevant if not perfect information. When initial nuclear power plants were designed and their safety was analyzed through a probabilistic risk analysis, some of their parts were novel but similar ones existed elsewhere. This was the case with nuclear reactors in the early submarines of the US Navy, which provided the best sources of experience that could be used when designing the first nuclear power plants.

- *Test data* They are critical, and their value depends on the test design. The more similar the parts tested to the actual part and their environment, the more valuable the test data. An example is that of shaking tables that permit testing the reliability of structures to seismic loads. The structure model has to be realistic—if not of the right size—and the seismic waves provided by the table must represent properly the loads to which the structure will be subjected.

- *Engineering models or social science models* They need to include accurate representations of both the loads and the capacities for the considered system. They depend on the nature and dynamics of the risk. One of the first detailed quantitative risk analyses was developed and used in the design and operation of nuclear power plants. After several iterations, they represented most of the uncertainties involved in the considered subsystems (US NRC 1975). The model of the accident scenarios includes first the initiating events. It could be, for example, a pipe rupture in the generator. Sub-models are then designed to assess the state of the plant after the sequence of events following an initiating event. In addition, external events can affect several subsystems and cause failure dependencies (Cornell 1980). It could be, for instance, whether there could be a failure of the core that causes radioactive release and loss of coolant. The next step is to assess the consequences of that accident. It depends on the quantity of radioactive material released, the weather conditions that can carry it, and the occupancy of the ground within its reach.

- *Expert opinions* They are both essential and controversial. Experts may have different levels of expertise, and some may not even know what they do not know. In addition, they can be biased and may try to influence the information and the decision to fit their own preferences. In general, the best use of experts is limited to parts of the analysis in which they have true experience and knowledge. Consulting several experts provides more complete information and may allow diminishing the effects of biases. Several medical doctors, for instance, may have different opinions about the risk of a specific disease given the symptoms of a particular patient. The problem is to aggregate expert opinions when they disagree. That can be done analytically by weighting the opinions of the different experts as a function of their credibility or simply choosing the judgment of one of them. More effectively, in some cases, the experts can be brought physically together around a table, so that they can discuss their sources of information and their mental models. It may be critical in medicine where disagreements may lead to different treatments and different outcomes for the patient.

A Long-Term Risk Analysis Is a Dynamic Exercise as New Events Occur and Information May Improve

The dynamic aspect of the risk analysis models is twofold. First, the frequency of accidents/incidents and of their nature vary and these variations can be captured by a quantitative analysis of these changes. Second, the knowledge itself of different system states may change—and hopefully improve—over time.

An important consideration is whether the risk is constant or not. As an example, following the tsunami that hit Japan in 2011 and caused an accident at the Fukushima nuclear power plant, a dynamic analysis of earthquakes and tsunamis in that area was published. There existed data since the year 840, showing an increase in the frequency of earthquakes of magnitude 8 or greater, such as that which caused the 2011 tsunami (Epstein 2011). A Bayesian analysis of these data allowed computation of the probability of such an accident, either assuming a constant rate of earthquakes or an increase in frequency as suggested by history. That type of information could be critical in siting, designing, or managing a plant in that area and elsewhere.

Again, one often does a risk analysis, static or dynamic, because there is not enough information about the considered system, but some can be gathered about the various subsystems and the different aspects of the risk can be brought together through a quantitative analysis.

The Format of the Results Includes Probabilities of the Failure Modes and the Distribution of Their Outcomes

The results can be of two different types: the marginal probability of an accident or system failure, or the probability distribution of the losses and damage caused by such accidents per time unit or operation.

In the first case, the result is simply the probability of a subsystem failure; for instance, the engine of a car. The second possibility is to assess the uncertainties about the different levels of consequences (mostly losses) of a system failure and to provide a probability distribution of the possible outcomes. An example, as mentioned earlier, is the probability of failure losses in earthquakes in a specific type of building and a given area (Cornell 1968). The result of the analysis is the probability that each loss level is exceeded, either in a specified time frame or per operation. This is the format of the risk analysis of nuclear power plants as currently done, which allows assessing the uncertain benefits of safety improvements in the different subsystems (US NRC 1975).

The Risk Analysis Method Can Be Applied to Technical Systems as Well as Human-Based Situations

Probabilistic risk analysis provides verifiable numbers that are the results of links and their uncertainties between scenarios and evidence. In other words, it is fact-based and should not involve preferences that are to be introduced into the decision phase of the analysis. As mentioned earlier, the method was developed in engineering to assess the failure risk of an accident in nuclear power plants. In the 1970s, these plants were new and complex, and the practical knowledge of their engines' reliability came essentially from reactors of US nuclear submarines.

The questions of risk and uncertainties encountered in engineering systems occur as well in social, medical, and political situations. The same quantitative risk analysis procedure is applicable based on key events and factors, their probabilities, their dependencies, and their consequences. For instance, assessing the risk of a new medical procedure requires gathering probabilistic information about an operation's procedure, the state and the physiological factors of the potential patients, as well as the skills of the medical teams. The problem of new systems occurs in that domain as well. It can be the case, for example, of the risk of a new medical device designed to clean a valve in the heart.

The same model can be applied in the political field—for example, to assess the risk of a terrorist attack in a given country (Bunn 2006; Kucik and Paté-Cornell 2012). This analysis requires an understanding as thorough as possible of the terrorists' scenarios and behaviors, the means at their disposal, their leadership, and the measures that they might consider when implementing their plan. In industry, for instance, when operating a system such as an offshore platform, one can assess the risk of cutting corners for lack of time, or due to the unavailability of parts that are necessary in a critical procedure. This was the case in the accident of the Piper Alpha oil rig in July 1988, which ended in explosions and fires as a young worker had failed to tag an inoperative pump at the end of a summer day (Paté-Cornell 1993).

When Several Actors Are Involved in a Conflict, the Risk Can Be Approached Through Game Analysis

Events and scenarios may involve several organizations or human beings with different and conflicting objectives. Game analysis allows assessing the probabilities and consequences

of decisions in situations of confrontation or competition (Paté-Cornell and Dillon 2006); for instance, the risks of insurgencies and terrorism (Paté-Cornell and Guikema 2002; Merrick and Parnell 2011). The quantification of these scenarios requires probabilities based on the knowledge of the strategies, the means and the preferences of the different adversaries. An example was the analysis of the risk of an insurrection in the Philippines, and of the moves of the insurgents against the government forces in the island of Mindanao. The risk was assessed considering alternative moves of both parties, and the effect of the different factors that motivated each party on the decisions of the other (Kucik and Paté-Cornell 2012).

Risk Management Is Best Supported by a Quantitative Approach to Risk Analysis

The bases of the quantitative risk analysis are systems analysis, scenarios of evolution including failure, and the probabilities of these scenarios and their outcomes. As such, they allow better and more accurate support of complex risk management decisions than simple guesses.

Quantitative information thus allows better risk management decisions than simple qualification generally does. First, it permits ranking risk management tasks by order of cost-effectiveness. In addition, it supports decisions that involve explicitly the risk attitude of a decision maker, who needs the quantitative description of the consequences of the different risk scenarios to estimate his or her utility for the outcomes.

Also, it allows judging if the residual risk as estimated probabilistically is tolerable given the uncertainties. The acceptability of a risk depends not only on its magnitude but also on the decision process. The risk magnitude and the uncertainties do matter in a reasonable analysis, and many organizations and governments have adopted quantitative thresholds and objectives. The question is how to know the chances that the objectives are met given the uncertainties. For example, a goal of less than 10^{-6} per person for a cancer caused by environmental pollutants has been set up by the Environmental Protection Agency in the US (US EPA 2005). Assessing such a risk requires knowing not only the amount of a poisonous substance to which someone can be exposed but also the dose-response relationship that determines the response of the individual. Given the diversity of the exposure and of the population, these factors are uncertain, but the mean of the result may allow for some credibility that the goal is satisfied.

Perhaps most importantly, quantitative risk analysis allows managing trade-offs of costs versus benefits at the margin—for instance, when deciding how much to allocate to different aspects of the risk factors. It could be that relaxing a constraint by one unit (for example, by allocating one more day or another million dollars to a particular task) would cause a significant risk reduction. Alternatively, it could be that the allocated risk management resources are too generous and ineffective at the margin, and that one could reduce the cost or the time spent without increasing the risk. In all cases, one needs the value of that risk to assess the shadow price of the constraint.

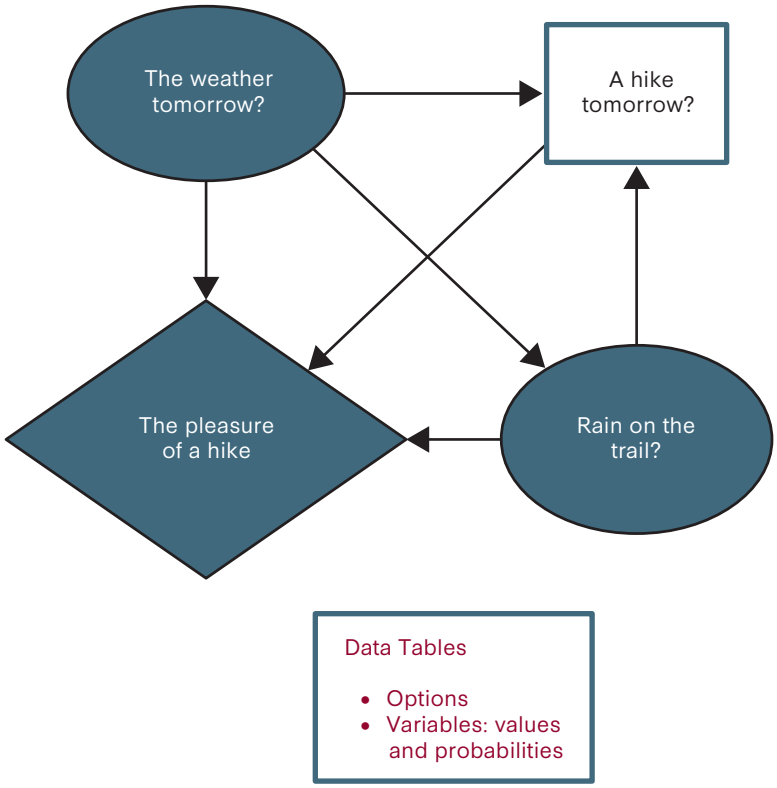
KEY GRAPHIC RISK ANALYSIS TOOLS: BAYESIAN NETWORKS AND INFLUENCE DIAGRAMS

INFLUENCE DIAGRAMS

To understand the role of key failures and events in the reliability of a system and its operation, one can use a graphical representation based on nodes and arrows called an influence diagram (Shachter 1988). Influence diagrams can be designed to yield the probability or probability distribution of outcomes, therefore a risk assessment result. They can also include decision variables, in which case they provide an optimal decision based on the maximization of an expected utility. The nodes then include these decision variables, classically represented in rectangles, random variables in ovals, and outcomes in trapezoids. The dependencies among them are represented by conditional probabilities shown by arrows between the nodes. Figure 1 shows a simple influence diagram example for the decision of taking a hike tomorrow given uncertain weather predictions.

An influence diagram is constructed in the following way: first, one draws the decision node (here, to go or not on a hike). Then, one draws the nodes that influence that decision; here,

FIGURE 1 An influence diagram for the decision to plan a hike tomorrow (no probabilities or outcome utilities, which are necessary in reality)



Source: Created by author for this paper

the uncertainty about the weather tomorrow, and about rain on the trail given the weather. The arrows among the nodes represent the conditional probability of their target given their source.

Note that in the influence diagram of the hike decision example, the utilities of the outcomes, the conditional probability distributions of the data—i.e., the marginal probabilities for the weather tomorrow—and the conditional probabilities for rain on the trail given the weather are encoded in separate tables.

A more complex and realistic quantitative example of an influence diagram is presented in the appendix.

Influence diagrams allow automatic resolution of decision problems in a quantitative setting. They are homomorphic to decision trees because they represent the same process and provide the same results, but they can be considered easier to construct and to use in communicating the analysis. In the case shown in figure 1, the weather prediction is the initial random variable, and its credibility influences the decision.

This formulation can be applied to all risk and decision analysis models—for instance, the risk of a terrorist attack. In that case, the diagram can represent the decision of the main decision maker (the defender) as well as that of the attacker to represent the game between the two and the links between the knowledge and the acts of both sides when making the next decision. The diagram representing a government's and a terrorist group's situations and decisions is presented in figure 2. The pointed red line shows the influence of insurgents' beliefs and actions on those of the government and vice versa.

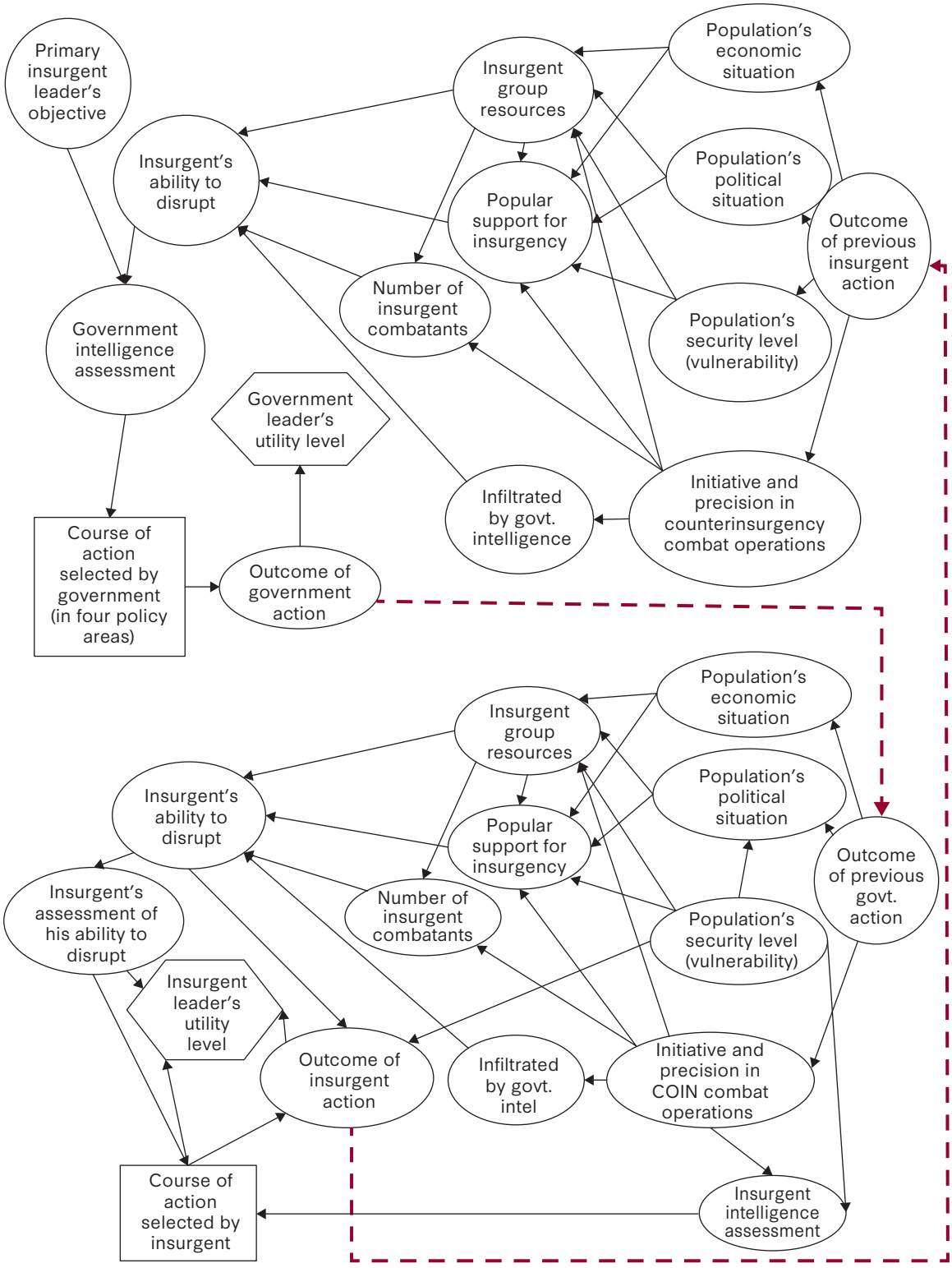
This representation is important because it shows the analysis of a two-sided game in a single diagram. That formulation could be expanded to additional players to address a game involving three or more sides. While it can be complicated, that analysis is feasible, contrary to a common belief that game models cannot represent more than two players.

Another example of an influence diagram is an application to the case of an oil tanker's loss of propulsion and the risk of losing oil (see figure 3). Key variables are the control of the drift and the location of the ship, either on the high sea or at a site where it can hit a rock, causing a breach in the oil tank. The measure of the risk is the probability distribution of the amount of oil that flows out of the breach and can reach the ground, and in addition, the damage that it causes.

STRUCTURE OF A RISK ANALYSIS MODEL FOR CYBERSECURITY

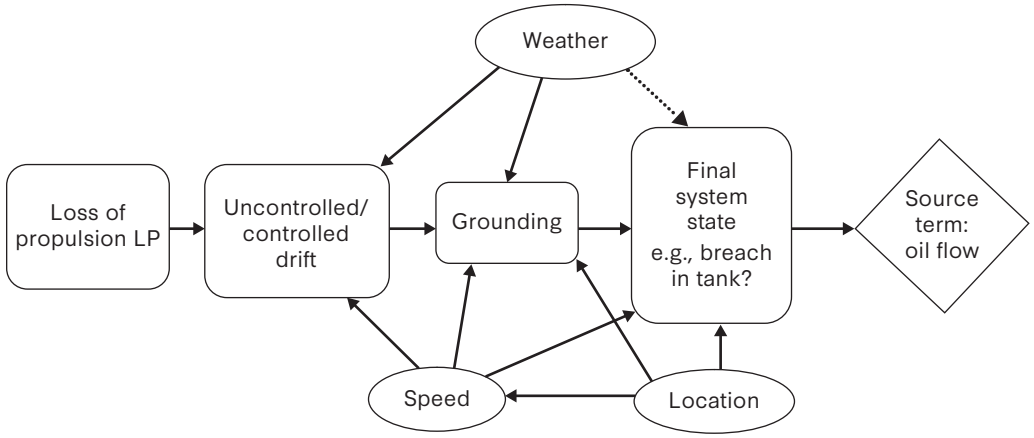
A probabilistic risk analysis using influence diagrams was developed and applied to several situations (Paté-Cornell et al. 2018). An example is the assessment of the risk of a cyberattack as shown in figure 4. Consider a target of attack such as a commercial company. A number of uncertain variables represent the factors of a cyberattack on that company. Several possible attackers may be set against it with different objectives (e.g., money versus technical

FIGURE 2 An influence diagram representing a confrontation game between a government and a terrorist group with information and decision links



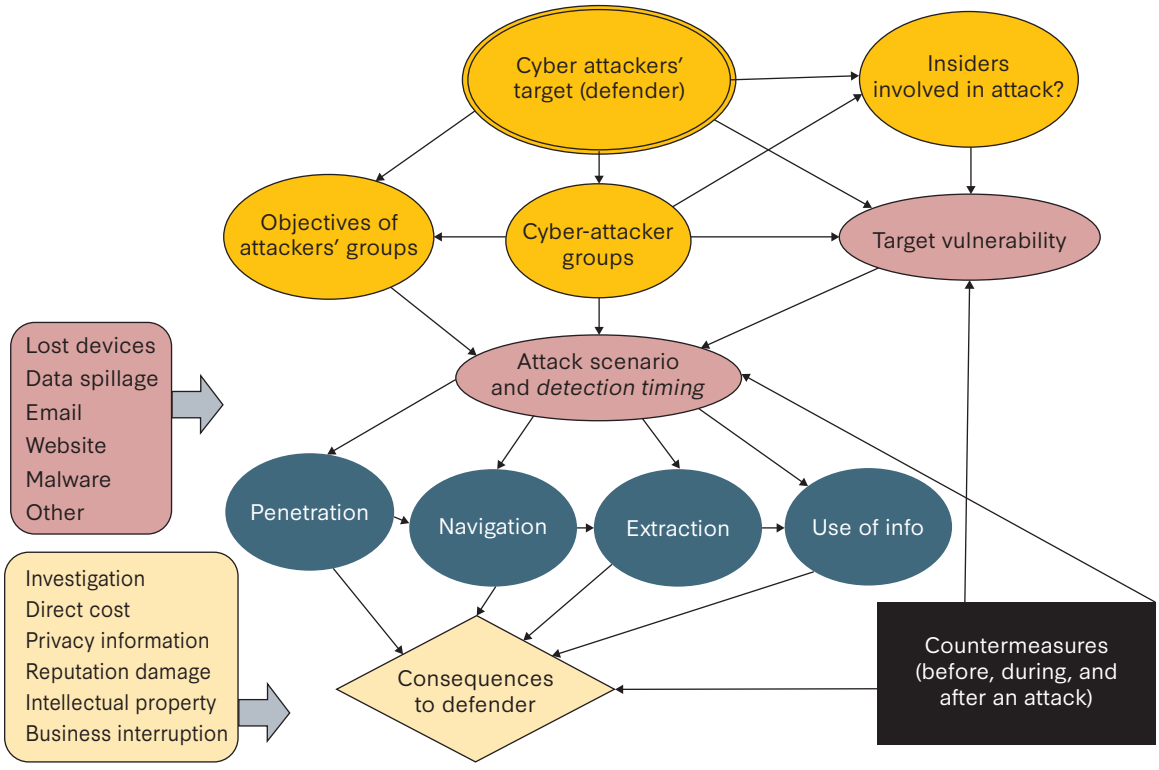
Source: Kucik and Paté-Cornell 2012

FIGURE 3 Influence diagram for a ship grounding example and the risk of oil spill



Source: Paté-Cornell 2009

FIGURE 4 Influence diagram representation of the structure of a cyber risk analysis



Source: Paté-Cornell et al. 2018

systems), and insiders of the organization may be part of the attack. Such an attack involves penetration of the computer system, navigation, extraction, and use of the information. The results are the consequences for the defender. They are often uncertain a priori. The choice of the attackers' target depends on the system's structure and the difficulty of penetrating it, the attackers' objective, whether insiders are part of the attack, and the effectiveness of the countermeasures taken by the organization before, during, or after the attack.

One critical aspect of the defense against a cyberattack is the timing of its detection. The attack scenario can be based on the use of lost or stolen devices, data spillage, emails, websites, malware, and so forth. The consequences to the defender of a successful attack can include the costs of investigation, the stolen device, intellectual property, loss of privacy, reputation damage, and business interruption. This model has proven quite useful, and was used, for instance, in the analysis of the risk of a cyberattack on the internal cyber structure of a space center (Paté-Cornell and Kuypers, 2023).

HUMAN FACTORS ARE MAJOR ELEMENTS OF SYSTEM FAILURE RISK ANALYSIS

The failure of technical systems often involves human errors. Yet, it should be noted that human interventions can have opposite effects; for instance, in some cases errors can cause a failure that would not have happened otherwise, or with a smaller probability. In other cases, human access and skills allow an operator to prevent a failure that would have occurred otherwise and caused severe damage. That was the case of the successful landing of US Airways flight 1540 on the Hudson River in 2009, thanks to the skills of the pilot.

THE MANAGEMENT ROOTS OF HUMAN ERRORS: THE SAM MODEL

A Critical Aspect of Human Errors Is That They Are Often Caused by Management Decisions

Managers decide who should be hired and how they should be trained. They also decide what should be done in specific crises, and with what incentives and constraints. An analysis of the risk can thus be done starting with management decisions, their effects on people's actions, and in turn, on the reliability of the considered system.

A more effective method is to start with the probability of failure of the technical system or the operations. The managerial risk can then be assessed based on three major steps: the analysis of the system (S), the operators' actions that affect each subsystem (A), and the management decisions that influence or determine these actions (M) (Murphy and Paté-Cornell 1996). That SAM model can be illustrated again by the case of the loss of ship propulsion, the control of the drift, and the collision with a rock unless the ship is at high sea. The structure of the model is as follows:

- *Step 1* The first step is the assessment of the probability of system failure as a function of the probability of failure of the different subsystems, their contributions to the overall

system function, their failure dependencies, and the role of external events. This analysis includes the identification of the failure modes (conjunctions of events leading to failure) and their probabilities.

- *Step 2* A critical part of that analysis includes the role of human operators and human errors. It requires identifying the operators of each subsystem and their decisions and actions that can cause a system failure. It may be a failure to solve observed problems, learn from a near miss, maintain parts of the system, identify a deficiency or a deterioration, and other human errors in specific parts of the system. To address the problem, it is important at that stage to try to understand the attackers' motivations. Again, operators' decisions and actions may involve some errors, but also positive moves that actually decrease potential losses and protect people.

- *Step 3* In turn, the management of the organization determines or affects the decisions and actions of the operators in charge of each of the subsystems, and thus the failure risk of the whole system (Paté-Cornell 1990). The managers hire the operators focusing on their competence, and they decide on their specific jobs and on their compensations given the quality of their performance. Key decisions at that level are the constraints of time and resources, which in turn may affect the quality of the work. Examples involve the maintenance of each subsystem and the frequency and depth of their inspection, which determine the reliability of these components including under external events.

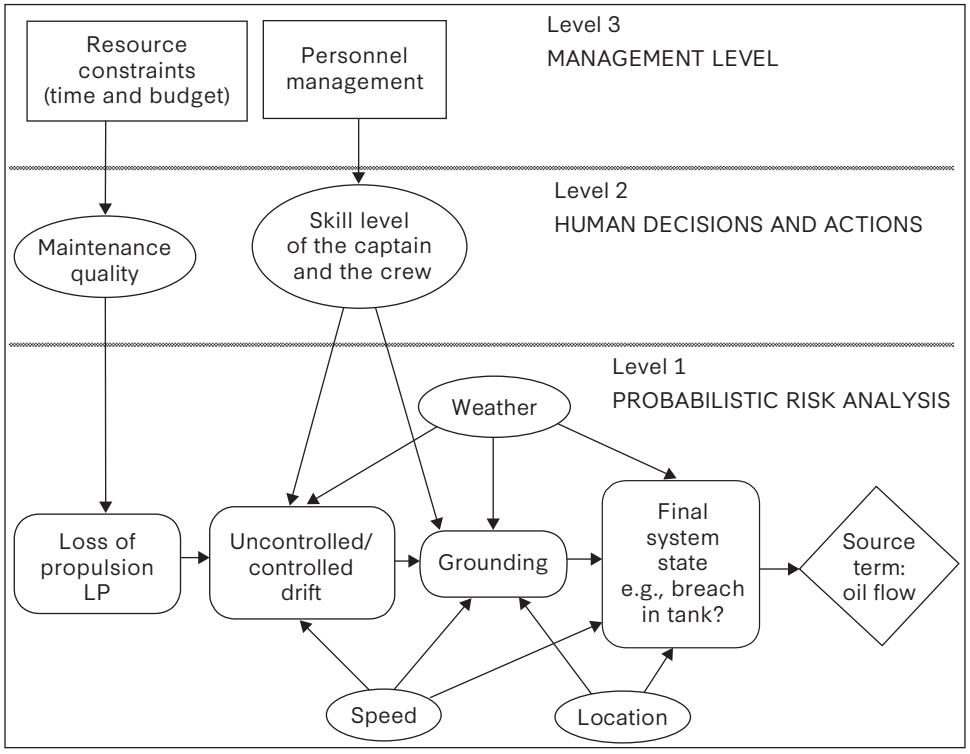
Explicit Linking of System Failure Risk to Human Decisions and Actions, and to Management Decisions

The global risk analysis model linking the systems' failures to management decisions includes first the risk of system loss (S), then the human decisions and actions (A) that affect that model's variables (especially the probability of subsystems' failures) and the management decisions (M) that affect these decisions and actions as represented in figure 5. That SAM model is most useful in guiding management decisions with the understanding of how they affect operators' behaviors and the system failure risk (Murphy and Paté-Cornell 1996).

The simplified example presented in figure 5 starts with the loss of propulsion of an oil tanker, which is linked to maintenance quality, and in turn to resource constraints of time and money. The control of the drift, which affects the chances of grounding and the energy of the shock, depends on the skills of the captain and the crew, thus by the management of personnel. This model allows getting to the root causes of accidents and supports risk management decisions better than a simple model of system failure and subsystems' performance.

An important feature of this SAM model is that it starts with the system's performance to identify the roles of the operators and in turn, the effect of management on system reliability. In reality, the causality chain starts with the management decisions that affect the operators' performance and thus the system's reliability. But the SAM structure is more effective in supporting risk analysis and effective risk management in different fields.

FIGURE 5 An example of the SAM model structure (System, Actions, Management) for the risk of an oil spill due to loss of tanker propulsion



Source: Paté-Cornell 2007a

RISK MANAGEMENT DECISIONS SUPPORT: RISK ANALYSIS AND DECISION ANALYSIS

A few basic notions regarding risk are involved in risk management decisions, as detailed below.

THE DECISION MAKER'S RISK ATTITUDE IS A MAIN RISK MANAGEMENT FACTOR

Risk management decisions are made under uncertainties about the facts that underline the risks. In addition, these decisions require explicit consideration of value judgments including a risk attitude (Abbas and Howard 2015). The values allocated to the different attributes that characterize the outcomes are those of the decision maker. These values may reflect dependencies in the combinations of attributes.

The risk attitude is an essential part of an outcome's valuation through a utility function, which shows the way the decision maker feels about the outcome. The risk attitude represents the variation of the utility (or disutility) with the different values of the attributes and their combinations. Considering potential losses, the preferences of a risk-averse decision maker involve

a greater increase in disutility with a specific variation of the losses at the higher end of the spectrum of outcomes than at the lower end. This implies that he or she fears an increment of high-level losses more than an equal increment of low-level losses. The reverse is the risk attitude of the risk-prone decision maker, who is more sensitive to a variation in the losses at the lower end than at the higher end of the loss spectrum. The risk-indifferent decision maker does not care about the starting point, and regards equally the variations in losses at any given point of the loss spectrum.

The risk attitude is key to decisions under uncertainties and depends only on the preferences of the decision maker. In other words, there is no right or wrong risk attitude.

THE FOCUS SHOULD BE ON THE DATA ONE NEEDS RATHER THAN THE DATA ONE HAS

Another aspect of the information relevant to a given decision is that one should not limit oneself to the data that are immediately available and easily gathered. What matters is to determine what data are needed, and to look for them as thoroughly as possible, recognizing that getting perfect information may be impossible and that some uncertainty is likely to remain.

Risk analysis and risk management information are generally imperfect and incomplete, which may be particularly critical in emergencies when there is no time to gather perfect information. A rational decision still has to be made under the remaining uncertainties, and the question is whether to gather additional data, given the costs and the value of that information.

THE ROLE OF THE DECISION MAKER IS ESSENTIAL, BOTH IN TERMS OF KNOWLEDGE AND OF PREFERENCES

Since decisions under uncertainty require the decision maker's value judgment and risk attitude, he or she should be identified if possible, when the decision is suggested by an AI system. This may not be feasible for a specific decision, and the analyst may have to consider the preferences of a group of people and enter in the algorithm what he or she thinks they may collectively feel about that risk. In that exercise, the analyst may lean toward risk aversion and want to be conservative—for instance, if the risk management decision will affect the protection of people.

THE QUEST FOR ADDITIONAL DATA SHOULD BE BASED ON THE VALUE OF THAT INFORMATION

Another critical issue is the choice of data, and more importantly, the judgment of what constitutes appropriate data given the limits of the existing dataset. It can be tempting to focus on what seems the most likely situations and to dismiss and ignore the rare extremes. It is an error that has caused failure to protect a system against rare events that proved catastrophic when they occurred.

One may thus need additional information, even though it may take time and resources to get it. Its value is determined by its effect—i.e., the improvement of risk management decisions. Importantly, the value of information depends on the risk attitude of the decision maker. The risk-averse, for instance, may consider more valuable the information that will permit a greater decrease in the probability of a severe outcome. Therefore, he or she will be willing to put larger resources into addressing the possibility of an extreme case.

An important point about the value of more data is that additional information does not necessarily decrease the uncertainties: one may discover a new scenario that was never envisioned and increases seriously the uncertainties about the outcomes.

ARTIFICIAL INTELLIGENCE IS CRITICAL BUT MAY INTRODUCE AN ALIGNMENT PROBLEM

AI may be an important part of risk management decisions since an AI system can provide relevant information and/or suggest options, given a larger database than provided by human knowledge. The information is as good as its sources and the processing of the data. Decision makers may trust the AI system if they believe that it is better than them at processing the information. But the decision itself depends in good part on preferences, and if they follow the opinion of AI, it will be dictated by the risk attitude that has been embedded in the system.

This implies that under uncertainties, the value of an AI decision algorithm depends on the alignment of its risk attitude with that of the human decision maker (Paté-Cornell 2023, 2025). Therefore, if the AI analyst knows who will use the system, he or she can include in the algorithm the relevant risk attitude. Most of the time, however, it is not the case, and the analyst may input into the system what seems a reasonable risk factor—possibly a conservative one out of prudence. The decision maker, if aware of the AI's preferences, has to decide whether to adopt its recommendation or to ignore it and make a decision consistent with his or her own risk attitude.

As described further, four examples of AI decision recommendations have been described in a previous publication (Paté-Cornell 2023, 2025): the choice of a medical test, a defense decision regarding the use of autonomous drones in combat, a sailing race where an AI system can give advice regarding a boat's trajectory and sail setting, and autonomous vehicles, in which the system makes risk management decisions that may fit the preferences of the population of riders. In all four examples, the main decision maker (the patient, the commander, the skipper, or a rider) has to think, explicitly or implicitly, of the chances of different possible outcomes and choose the optimal option—either the choice of the AI system, or that based on his or her risk attitude. The AI system thus needs to be aligned when there is a discrepancy between its preferences and those of the decision maker.

Aligning the AI system, if needed, thus requires that the algorithm be accessible, that its risk attitude be explicit, and that the system be set for the encoding of the actual decision factors.

EXAMPLE 1: PATIENT RISK IN ANESTHESIA

THE SYSTEM INVOLVES BOTH THE PATIENT AND THE ANESTHESIA TEAM

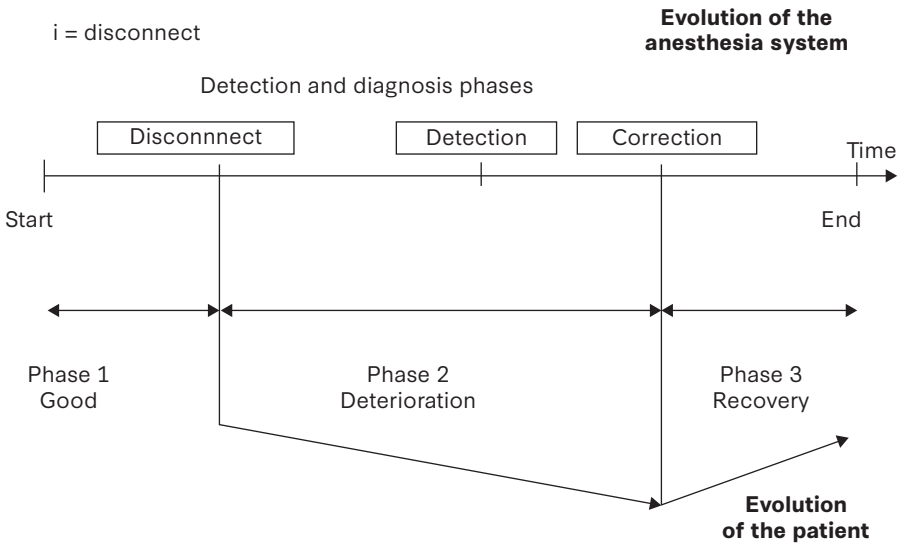
In this risk analysis study, the system included both the surgery patient and the anesthesia team. The surgical procedure was a safe one, such as knee surgery, and the risk of the surgery was limited to the anesthesia (Paté-Cornell et al. 1997; Paté-Cornell 1999).

A key factor regarding the patient is his or her resistance to surgery and anesthesia (see figure 6). For instance, some patients such as obese human beings and premature babies are more vulnerable to anesthesia. The other part of the system is the anesthesiologist, whose performance is affected by his or her competence and alertness.

Competence may be an issue, in particular with residents who may not yet have the experience that it will take to face a serious problem. Diminished alertness may also be an issue for any practitioner (for instance, due to lack of sleep) when a patient problem needs to be detected, understood, and corrected.

The patient's case may be a complex one that will require a high level of anesthetist competence.[3] One must acknowledge the role of the nurse-anesthetists in the operating room. Some of them have experience that may allow them to supplement that of the anesthesiologists, and detect, for instance, tube disconnect problems that need to be addressed on the spot.

FIGURE 6 The development of an anesthesia accident



Source: Paté-Cornell et al. 1997

Furthermore, the rules and the environment of the hospital may affect the roles and the respective responsibilities of the surgeon and the anesthetist. In critical decisions, in particular, a clear allocation of duties may affect the safety of the patient.

CRITICAL INFORMATION AND INFORMATION SOURCES

A hospital in Adelaide, Australia, was the source of part of the information used in this study. In the United States as in Australia, the risk of a severe anesthesia accident in an operation is in the order of 1/10,000. What is considered here a severe accident is brain damage or death. Bayesian probability was used to assess the probabilities of the different events in accident sequences, starting with initial events such as the disconnection of the oxygen tube linked to the patient's lungs, or a mistake in intubation in which the tube was inserted in the stomach.

To analyze the unfolding of an accident and short on statistics, one needed expert opinions. Those came from anesthesiologists, active or retired, but also from surgeons and from nurses who were particularly helpful.

Focusing on the dynamics of an accident and the way it was handled, detection and medical reaction times were key risk factors. These included the time that it took to observe an incident after it happened, to identify the cause, and to address the problem. Next, of course, the question was whether the detection and reaction were correct, and the problem was solved.

The dynamics of an accident are thus critical to the risk analysis and can be addressed through a stochastic process.

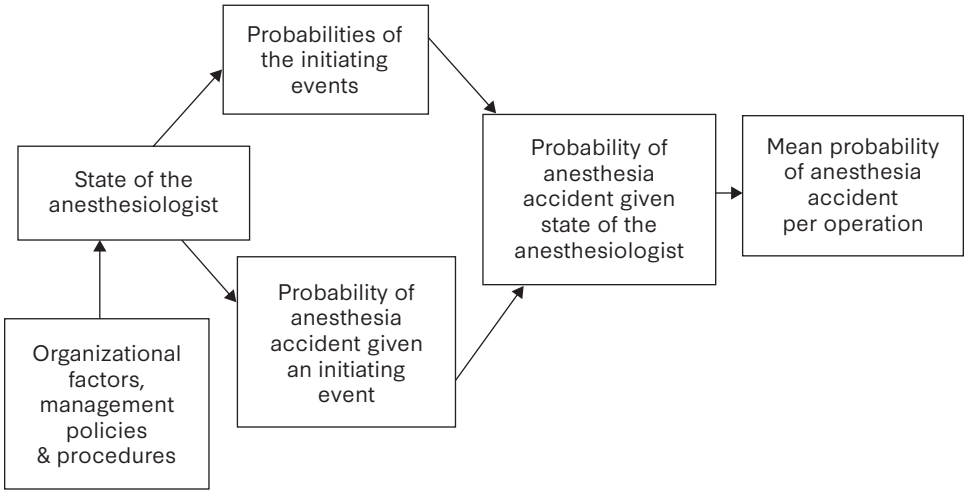
THE DYNAMIC MODEL OF ANESTHESIA RISK MANAGEMENT

A dynamic model of risk analysis includes quantification of the duration of events in accident sequences, and of the chances of the possible outcomes.

A general anesthesia accident sequence involves first the occurrence of an initiating event such as a drug allergy (see figure 6). Next is the reaction of the practitioner, who may take a few minutes to observe and understand the problem. Another issue is the time that it takes for the patient to react to the drug and the medical intervention, and hopefully to recover. But this is a case in which the patient's tolerance for a particular drug and the time he or she can live with an allergy—if they have any—affect the risk of an accident, possibly a deadly one. The question here is thus how the risk is influenced by the competence and alertness of the practitioner, and by the sensitivity of the patient to a specific drug.

The hospital's management decisions may involve education, hiring, recertification, and the rules of interaction. The reactions of the practitioners depend in part on these factors, and how they affect the performance of the anesthetists. But a key issue is their ability to face a crisis given their personality, the training that they have received, and the monitoring of their behaviors.

FIGURE 7 The structure of the anesthesia patient risk model



Source: Paté-Cornell 1999

THE STRUCTURE OF THE QUANTITATIVE RISK ANALYSIS MODEL

Figure 7 represents the structure of the risk analysis model, focusing on the anesthetist and the organizational measures that affect his or her performance.

As mentioned earlier, statistical data were available for the initiating events and the overall accident probability. The first element of a scenario is the initiating event, which triggers a subsequent set of events that may lead to an accident. The analysis allowed assessing the risk for each type of possible accident in an operation, then the global risk. One of the initiating events is the disconnection of the tube linking the oxygen supply to the throat of the patient if the anesthesia is administered by ventilation. Other initiating events include esophageal intubation, non-ventilation, malignant hyperthermia, inhaled anesthetic overdose, serious allergic reaction to the anesthetic, and severe hemorrhage. Each of them may be followed by an accident sequence, which depends in large part on the reaction of the anesthesia team and on the state of the patient.

Detecting the evolution of accidents then depends on observing and reacting to signals. The first issue is whether the problems are visible and within what time they occur. For instance, short of oxygen the patient may turn blue, and the problem can be diagnosed, hopefully in time for correction. The success of the anesthesia team in detecting and responding to signals thus determines in large part the outcomes of initiating events. The probability of these outcomes is the product of the probability of the initiating event and of the conditional probability of each scenario that may follow.

The probability of an accident given an initiating event is thus the result of a dynamic model of accident sequences. Assuming that the initiating events could be the result of a human error, the risk depends mostly on the performance of the anesthetist. One can then assess the probability of an accident for each initiating event, then for a given operation.

Structure of the Risk Model Given Human Response and Management

For each scenario, the effect of the anesthetist's performance on the patient's risk is computed by considering its role in the occurrence of an initiating event, and in the possible subsequent events that may lead to an accident. The patient risk of death or brain damage can then be linked to the performance of the anesthesiologist using the following structure and considering the following factors.

- *The organizational factors* For example, the time the practitioner has been on duty may affect his or her performance. Anesthesiologists generally cover the full operation, and the time on duty may reach as long as 12 hours.

- *The state of the anesthesiologist in terms of mood and ability* It also affects his or her ability to perform without problem. It may be a characteristic of the personality but may also depend on the case and the time constraints.

- *The probability of some initiating events of anesthesia accident* These events can lead to death or brain damage. The problem may depend on the performance of the surgeon (for example, a hemorrhage) but the outcome is a function of the reaction of the anesthesiologist, whose role is to control the hemorrhage and manage the blood supply at his or her disposal. The risk to the patient is thus a function of the intervention, and to some extent, the state of the anesthetist.

The analysis allows linking the competence and behavior of the anesthetist to the time it takes him or her to react, and to the state of the patient. Given the adequacy of that intervention, one can assess the benefits to the patient of organizational improvements in the practitioner's work environment.

EXAMPLES OF MANAGEMENT MEASURES THAT AFFECT THE ANESTHETIST'S PERFORMANCE

The state of anesthesiologists in terms of competence and alertness is affected by their personality and by the training and circumstances that determine their ability to face problems in the operating room. An important part of their training is done on simulators.

It is also a function of the surgical operation and of the necessary equipment. A number of measures that affect these factors can be taken by the hospital, and by the profession in general. The work schedule depends on the frequency and the nature of operations. The anesthetist function starts at the onset by putting the patient under anesthesia, after which he or she needs to be monitored, sometimes for hours. Therefore, the anesthetist needs to be paying attention, even in times when things go smoothly and the patient seems to be doing fine. That phase of the work has sometimes been compared to that of a copilot who may feel that he should not need to intervene. In fact, given that kind of perception, some anesthetists have been observed doing something else, even briefly leaving the operating room.

Some factors that affect the state of the practitioners include the following:

- *Selection and periodic training of anesthetists* All medical students may not have the right personality to be anesthetists, and in particular to face critical situations that may lead to an accident and require immediate and intense attention. Their initial training is thus critical, but so is their continuous education through experience, sharing situations faced by their colleagues, and understanding the way in which they were resolved.

- *Supervision of residents* A critical part of that training is the supervision of residents. They are doctors already but may not have the experience that allows them to face serious and/or rare situations. Even later in their career, some anesthetists may not be able to handle problems that they have never encountered before or have forgotten, and they may need some reeducation or a backup who can take over.

- *Equipment performance* Equipment does not appear to play a critical role in the risk of an anesthesia accident. Nonetheless, inspection and maintenance is an important part of the process. For example, oxygen needs to be available, and the conducting tube needs to be functional and plugged.

In fact, these factors are general problems in many industries and can be handled by rules, regulations, incentives, and rewards.

EFFECT OF A MANAGEMENT POLICY CHANGE ON PATIENT RISK: EXAMPLE OF SIMULATOR TRAINING FOR ANESTHETISTS

Consider the policy of one day of simulator training per year, for experienced anesthetists who may lack regular training and may have forgotten the adequate response to some situations. On the one hand, this measure may be burdensome to some who do not believe that they need it. On the other hand, it will allow others to remember procedures that they may have forgotten, or to get familiar with new technologies or situations to which they have never been exposed.

PATIENT RISK REDUCTION: ANALYTICAL RESULT

If the measure of retraining experienced practitioners is enforced, they will encounter rare events, first on a simulator and perhaps later on a patient. The model described earlier was applied to that case, using statistics and expert opinions. It was found that this policy change may reduce the overall risk of patient accidents by 16 percent, assuming that the problem is fixed; i.e., that the experienced anesthetists actually learn about procedures that they have forgotten or never encountered before.

CONCLUSIONS OF THE ANESTHESIA STUDY AND THE EFFECTS OF CULTURAL FACTORS

Drug and alcohol abuse, which were the initial motivations of the study, were not found to be major contributors to patients' risk. This does not mean, of course, that rare problems should

not be given serious attention, but the major contributors to patient risk attributable to anesthesia were routine problems such as distraction or incompetence of the practitioner.

The main procedural improvements and the best risk reduction benefits were found to be the following:

- *Formal recertification of practitioners* It was found to reduce the patient's risk of an accident by 23 to 29 percent per operation.
- *Regular simulator training of the practitioners* Such training reduced the patient's risk by 16 percent.
- *Improving the supervision of residents* Improved supervision reduced the patient's risk by 14 percent.
- *And to a lesser extent, limiting anesthetists' time on duty* This measure reduced the patient's risk by 6 percent.

What is important in this case is that the risk analysis showed that what the hospitals were most concerned about—i.e., drug and alcohol abuse—were not the major contributors to patients' risk. Even though these abuses had received a lot of attention, they were not in fact the most dangerous. Instead, it was more routine events that may not be as visible or noteworthy yet needed to be addressed even when they were problems of regular, sometimes well-regarded, practitioners.

This kind of surprise is in fact often the result of risk analyses because people do not pay as much attention to common problems as to rare ones, and risk management efforts may be focused on more visible rather than more relevant events.

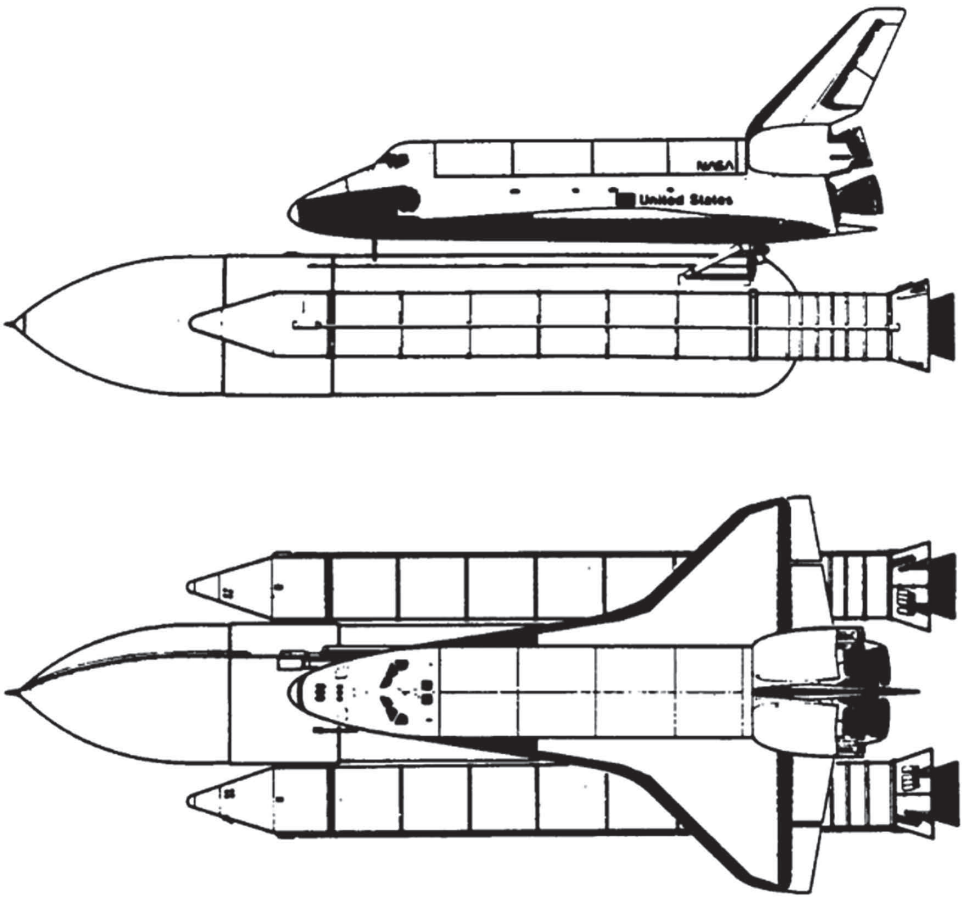
A TECHNICAL RISK ANALYSIS EXAMPLE WITH ORGANIZATIONAL ROOTS: THE SHIELDS OF SPACE SYSTEMS

THE ROLE OF HEAT AND RADIATION SHIELDS

The US space shuttle needed to be protected against heat at reentry into the atmosphere. Other systems, such as the Europa Clipper that orbits around Jupiter, need to be protected against radiation (Ding et al. 2020). The questions are: What is the risk of losing a system through a failure of a given shield, and how to provide adequate shielding?

The issue is one of load; that is, the heat or the radiation versus the capacity of the shield and of what it protects, and therefore what load the system can tolerate without failure. The heat shield may sustain different kinds of heat loads that could create a gap in the skin of a spacecraft. If that happens, the hot gases penetrate the surface and could destroy critical subsystems under the skin.

FIGURE 8 The space shuttle assembly: the orbiter, main tank, and two solid rocket boosters



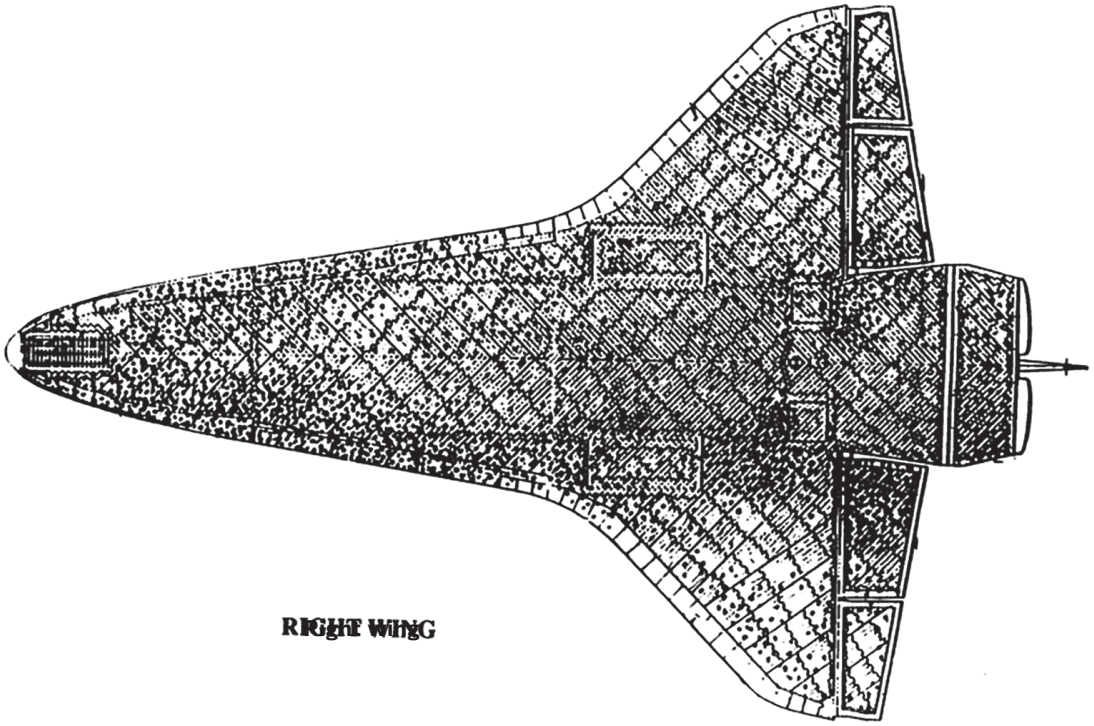
Source: NASA

THE TECHNICAL FAILURE RISK OF THE HEAT SHIELD

A risk analysis was performed for the US space shuttle heat shield before the program was terminated in 2011 (Paté-Cornell and Fischbeck 1993). The shuttle system consisted of the orbiter, attached to a large tank containing liquid oxygen and liquid hydrogen, and two solid-fuel rocket boosters that assisted the main engines at takeoff (see figures 8 and 9). The orbiter was protected by about 25,000 black tiles glued to its aluminum surface through a felt pad. The tiles could debond at takeoff if they were not properly attached to the orbiter skin, or if they were hit either by pieces of the shield of the external tank (as was the case with the space shuttle orbiter *Columbia* in 2003) or by other kinds of debris such as ice, stones, birds, and so forth.

The risk analysis was based on maps of the space shuttle orbiter representing zones characterized by the values of four key factors at each point of the orbiter: the heat load, the debris density, the aerodynamic forces that may cause the loss of adjacent tiles if one is detached, and the sensitivity to heat loads of the different parts of the aluminum skin and subsystems underneath. For each factor, a map of the orbiter showed its intensity in different zones.

FIGURE 9 The black tiles of the heat shield of the space shuttle orbiter: cumulative hit load (debris density) during about 30 flights



Source: NASA

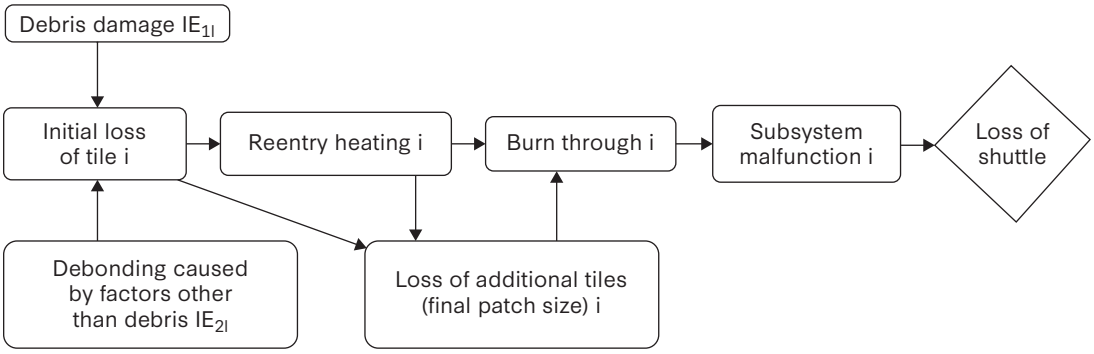
The structure of the risk analysis model is shown in figure 10, which is an influence diagram showing the uncertain events and variables from the initial loss of tiles to the gaps in the orbiter skin, their size and their effect on the subsystems exposed to hot gases, and finally, the potential loss of a mission.

The result of the analysis was a map of the orbiter's tiles showing the risk contribution of each zone to mission failure (figure 11). In that figure, the darker the area, the more risk-critical the tiles in that zone. That map was communicated to different space centers, and in particular to the tile maintenance team at Kennedy Space Center, where it was used. The contribution of the tiles to the failure risk was shown to be about 1/1000 per mission, i.e., about 10 percent of the failure risk of a mission, which was in the order of 1/100.

ORGANIZATIONAL FACTORS OF THE HEAT SHIELD FAILURE RISK AND RISK MANAGEMENT

An important part of the study was the extension of the technical failure risk to the organizational factors that contributed to it. A number of organizational improvements in the treatment of the tiles were considered. They included relaxing the schedule constraints of the tile maintenance between flights, testing the most risk-critical areas by hand before takeoff, and improving the heat shield of the external tank to reduce the risk that pieces of it could debond and hit the black tiles of the orbiter.

FIGURE 10 Influence diagram showing the analysis of the risk of losing a space shuttle mission due to a failure of the heat shield tiles



Source: Paté-Cornell and Fischbeck 1993

Some of these measures were implemented but not the improvement of the attachment of the heat shield to the external tank. During the last *Columbia* flight in 2003, a piece of it detached, hit the left wing, and caused a large gap that allowed hot gas to destroy critical parts of the orbiter. The space shuttle orbiter exploded, killing all astronauts on board.

The study had been made publicly available before the accident but was rejected by one reviewer because it was not based on "data," explicitly restricted in his mind to mission failure statistics due to the tiles. Since the accident had not happened yet, as is sometimes the case, there were no such data available. The lesson was learned, however—further risk analyses of the shuttle were performed and allowed improvement of shuttle subsystems.

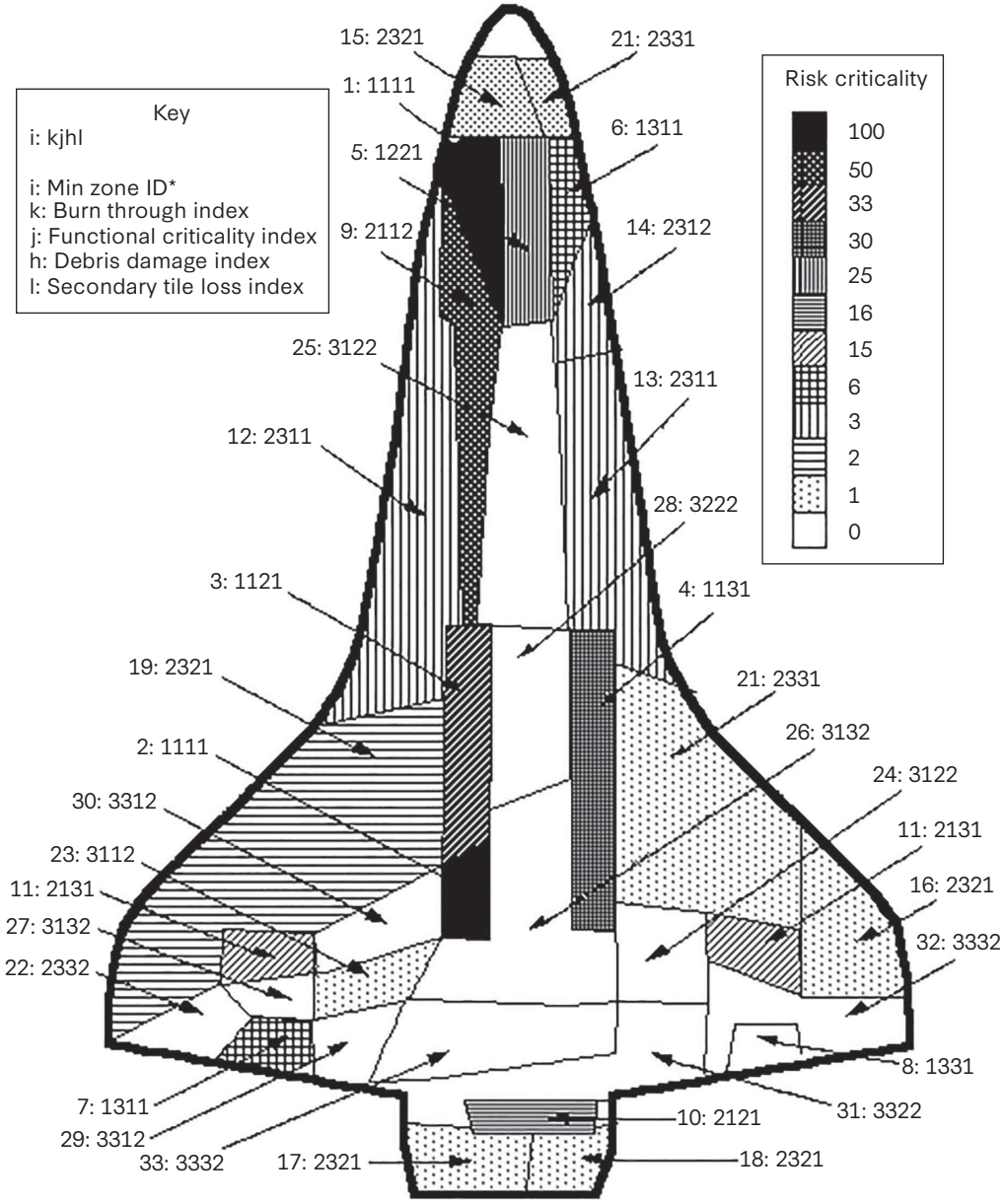
The lesson was that, of course, one does not need the failure of an engineering system to justify a risk analysis that will allow global improvements. One needs a set of data characterizing the performance of the different subsystems and their dependencies to be able to assess the global failure risk before (enough) statistics are gathered.

ARTIFICIAL INTELLIGENCE AND RISK MANAGEMENT: THE PROBLEM OF AI SYSTEMS ALIGNMENT

AI SYSTEMS PROVIDE TWO THINGS: INFORMATION AND DECISION SUPPORT

Information can feed the analysis of failure risk, risk management options, and risk reduction benefits. Some of that information is deterministic, some probabilistic. Some of that information is comprised in an initial database, trained then enriched by machine learning as experience is gathered, including elements of the model described earlier, such as failure modes and effects of external events. Note that the information may include errors and biases; it may be limited, but in general provides some notions that may not have been gathered by human experts.

FIGURE 11 Results of the risk analysis: contribution of the tiles in different minimal zones to the loss of a shuttle mission



Source: Paté-Cornell and Fischbeck 1993

AI RISK MANAGEMENT DECISIONS

Decisions under uncertainties are supported by the algorithm based on that information, but also by a risk attitude encoded in the program. This is an important feature that is not always understood by the people who receive the message—i.e., the managers and users of the system—who do not question the AI system given the span of information that it relies upon. The risk attitude is just one factor embedded in the algorithm but not explicit to the user.

Yet, whereas the information can be justified and corrected if needed, there is no correct risk attitude, except if one believes that the risk preference of a crowd or an organization should prevail. The choice and the source of the risk attitude are thus critical to the relevance of the AI message.

That choice may pose an alignment problem if a specific decision maker, responsible for risk management, wants his or her risk attitude to prevail (Paté-Cornell 2023, 2025). For instance, based on the diagnosis of an AI system, the owner of a small plane may decide that what seems to be a minor problem is too dangerous for a flight. Another pilot may decide that the risk is tolerable. In any case, the AI system provides a judgment and a decision, based on a risk attitude implemented in the algorithm that may or may not be that of the actual decision maker.

In this general alignment problem, the AI parameters represent preferences that need to fit those of the decision maker and/or the organization that manages the system.

THE AI ALGORITHM AND ITS OUTPUT

All risk management decisions under uncertainties thus reflect a risk attitude. Yet, as mentioned earlier, there is no "right" risk attitude, unless it has been explicitly specified by the risk management organization. For instance, most deficiencies, small as they may be, in nuclear power plants or in an airplane, require attention. But the management of small defects in a car depends on the risk attitude and on the resources of the owner.

If the risk attitude is not specified elsewhere (e.g., by polls among the population regarding a specific risk), the risk attitude of the AI decision algorithm is the analyst's choice based on opinions of experts or a crowd, of public intuitions and fears, or other factors.

If the algorithm includes a utility, the risk attitude is included in that function. Also included in the utility are the values that the decision maker chose for each possible outcome for each attribute. That utility is continuous and monotonic (it grows with the outcome value) and the risk attitude can be assessed from that function.[4] This is important because if one wants to choose the optimal option given the preferences of the decision maker, the AI system needs to be aligned to these preferences. To do so, one may need to "unzip" the algorithm, extract from it the value of that risk attitude, check whether it fits the decision, and modify it if it does not.

To ensure the alignment, one needs to assess the risk attitude of the decision maker by presenting him or her with "lotteries"; i.e., choices based on probabilities and values of outcomes, and ask what would be, for example, the "selling price" of that lottery that he or she would be willing to pay to get rid of the uncertainties. By equating the utility of that selling price (the "certain equivalent" of the lottery) to the expected utility of the lottery, one may be able to assess the utility of a sufficient number of outcomes, and therefore a risk attitude. That often implies drawing a utility function through the points that one has gathered, in order to assess an estimate of the risk attitude as the ratio of the second derivative to the first (minus), at least in the relevant outcome segments.

THE HUMAN IN THE LOOP PROBLEM

The question is whether there is a human in the loop at decision time, or whether the AI decision is automatically implemented. If the best option is determined by AI alone, but the decision itself is made by someone else, there may be an alignment problem with the risk attitudes. One cannot make the comparison unless one gets to know the attitudes of both the AI system (which is not obvious) and the actual decision maker. As will be shown in some examples further, that issue is critical—for example, in medical choices or national security decisions.

AN AI SYSTEM MAY NOT BE TRANSPARENT: WHAT IS ITS RISK ATTITUDE?

If a human makes the decision based on information provided by an AI system, it is important for the decision maker to understand the source of the AI opinion, both in terms of information and preferences, and to know the system's risk attitude.

CAN THE DECISION MAKER OVERRIDE THE AI DECISION?

In some cases, the decision is automatically implemented. In others, the AI system provides the option that the algorithm identifies as optimal, which reflects the risk attitude embedded in the program. It could be that the AI system is more risk-averse than the decision maker or that they do not use the same trade-offs among attributes. The question is whether the AI system fits the decision maker's preferences.

For example, in a medical case, the patient may want to avoid a painful procedure, even if the judgment of the AI system is that it is the safest choice for the global population. These trade-offs, however, will be felt differently by different people, and the objective for the patient is to choose the option that he or she prefers given the uncertainties.

FOUR REAL-LIFE EXAMPLES OF AI RISK ALIGNMENT PROBLEMS

The alignment problem and the use of AI systems as providers of decision guidance has become essential in many domains and will be increasingly so as large databases and large language-learning models will be implemented in a number of fields of life.

What is presented here is a set of four examples that are realistic but illustrative (Paté-Cornell 2025). It is assumed that in each case, a human being will have the choice between two options: following the AI system's recommendation, thus accepting in some cases an automated decision, or using his or her own judgment, knowledge, and risk attitude to make a decision that may be different from that of the algorithm.

A Medical Case

A patient had chosen to get a mammogram without any symptom of breast cancer because it was a CDC recommendation that the test be performed given her age. There was a minor image problem in the visual equipment, and her doctor recommended a needle test, which came back negative. Nonetheless, a risk-averse AI system recommended that she get an

additional surgical biopsy. The patient had to decide whether to follow that advice. In this case, given the very low probability of a disease and the downsides of surgery, and with the advice of a higher-level practitioner, she did not accept the test, and decades later has never had breast cancer.

A Defense Case

Automatic drones are used to hit specified targets with a human in the loop. By contrast, autonomous drones guided by an AI system choose their target and can respond automatically and immediately to an enemy attack. The use of autonomous drones in combat is, for the moment, forbidden by the US Department of Defense. The US and China (as far as we know) have some of these drones but have not used them in combat. Other countries such as Russia and Turkey do use them, and the US may have to consider the speed of ballistic exchange and the use of autonomous drones in combat to avoid the immediate destruction of US capabilities. The problem with that option is that it may not allow a commander to stop an exchange of weapons to allow for negotiation of a truce that could decrease the losses on both sides.

A Sports Case: Sailing Races

A sophisticated AI system was at the core of the victory of the New Zealand team in the 2024 America's Cup. Yet, there are generally remaining uncertainties in the AI evaluation of the winds, the currents, and the performance of the competitors. Most AI sailing guides are not as sophisticated as the New Zealand one but still provide valuable information. Given an AI evaluation of winds and currents, should a skipper change her course or the setting of her sails? It depends on the sophistication and the accuracy of the AI system and on the experience of the skipper. In the end, it is her decision given her knowledge of the area, her understanding of the competition, and her actual objective in the race (try with a small chance to finish first or ensure a decent position even if it is not among the top ones). In any case, one key factor in that decision is her risk attitude, which may or may not be that of the AI system.

Autonomous Vehicles

Autonomous vehicles are generally safer than other vehicles, yet susceptible to accidents, many of which are caused by other cars (e.g., rear-ending). They function without information about a rider's risk acceptance given interactions with other vehicles. The autonomous vehicle's reaction is dictated by the AI system and thus depends on the judgment introduced by the analyst. In the design of the algorithm, risk attitude is part of the vehicle's management software. The analyst's opinion becomes the risk attitude of the passenger, which varies by definition and is part of the algorithm. One could imagine giving the passenger the option to set automatic risk control; but first, that would require on the passenger's part an understanding of his or her risk attitude, and second, it could trigger a legal challenge.

In all four cases, the ultimate choice should depend on the risk attitudes of the decision maker and of the AI system if they can be aligned. Otherwise, it may only be the judgment of the analyst, thus reflecting a risk attitude that is embedded in the AI system without an explicit link to the actual decision maker.

SOME APPROACHES TO SOLUTIONS OF THE AI ALIGNMENT PROBLEM

Aligning the risk attitudes of the AI system to that of the future decision maker(s) is generally needed in the decision stage if AI is playing a decision role. This adjustment involves several aspects of the AI system.

- *Revealing explicitly the sources of AI information* This is one essential way to allow the decision maker to assess the validity of the probabilistic information.

- *Flexibility of the AI system* One needs to permit adjustment of the risk tolerance factor either by allowing "unzipping" of the system to modify that factor in the algorithm, or by allowing the decision maker to input his or her risk attitude in the software before making the decision. In both cases, this implies access to the risk attitude of the software and possible modification of it, while preventing adversaries from accessing and modifying the system.

- *Education about preferences* The decision maker must be able to assess his or her own risk attitude in terms compatible with those of the AI system, and to check it against the AI's. To do so, she or he can assess and compare lotteries as is shown further, and the risk attitude can be derived from evaluation and comparison of these lotteries.

- *Communications, when feasible, between the analyst and the decision maker* Communications may be essential between the analyst and the user if the analyst is designing the algorithm to guide a specific, known decision maker in the decisions to be made. The problem is to know who decides and what risk he or she is willing to take. But the solution should also involve the people who will have to live with the consequences of the chosen option—for instance, medical AI recommendations.

AI-BASED CYBER RISK MANAGEMENT: AN EXAMPLE OF WARNINGS OF A CYBERATTACK

The assessment of warnings of cyberattacks to guide the response of the defender is an example of the general risk analysis model presented earlier. The model presented here is based on a specific but illustrative case of cyber defense that demonstrates a general approach to cyber risk management supported by an AI system (Faber and Paté-Cornell 2020).[5]

AN ILLUSTRATIVE SITUATION AND ITS ACTORS

- *A commercial corporation and its cyber system* A company, for example selling t-shirts, wants to protect its main computer, which contains data (commercial, financial, technical, etc.) that some attackers would like to get.

- *Two kinds of actors: customers and attackers* Both are trying to enter the computer system. The customers need to place their order, and cyber attackers are trying to reach its core to access and steal the information that they want to get. In the study, the behaviors of the attackers were observed through 18 "honeypots," which are simulated targets that were set up to attract and identify potential attackers.

- *A hybrid defense system protects the computer* It involves both a human expert and a robot. Therefore, both an AI system and a human being are in the loop.

- *Opening or closing a gate* The defenders' decisions are to open or close the "gates" in the computer system when actors are observed progressing through them. A gate is understood here as one step in the kill chain (the sequence of attack functions) of the cyber attackers, but the customers go through some of the same "gates" even though they do not intend to attack the system.

- *Who makes the gate decisions* In the beginning the human decides whether to open or close these gates. The robot observes and learns.

- *Robot learning* The robot (the AI system) learns from the human expert who makes the original sequential decisions, then takes over the control of each gate when its level of competence is deemed sufficient by the algorithm.

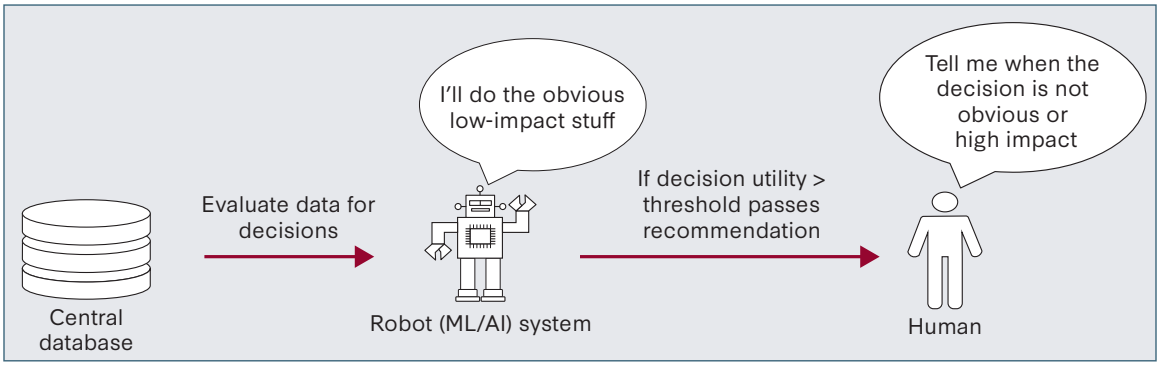
- *The expert can take back control* When the level of uncertainty or the size of the outcomes is too high, the robot passes the hand back to the human, who then decides whether to open or close the gate before returning control to the robot.

The hybrid cyber defense system is represented in figure 12, showing the interaction between the two defense actors. The defense team observes and follows each entry into the computer. At each gate, it makes the decision to keep it open or to close it, understanding the possibility that the actor is an attacker but that it could also be a legitimate customer. That decision is a rational one, based on probabilities and an expected utility, thus on the uncertainties about the actors and the value of the outcomes in both cases (a loss or a sale). That utility is a value function assumed to be that of the decision maker or the organization, and to include the corresponding risk attitude.

The defenders thus must make gate decisions under uncertainty given a trade-off based on the probability that the actor is an attacker or a customer. The value function that guides the defenders' decisions to open or close each gate is the utility of the outcomes to the defenders—a sale if the actor is a customer, a loss if the actor is an attacker.

As shown in figure 12, the robot passes the hand to the human when the threshold of uncertainties or outcome size exceeds a level specified in the algorithm.

FIGURE 12 The hybrid cyber defense system



Source: Faber 2019

MODEL OF THE DEFENDERS' DECISION SEQUENCE

One can thus describe the cyber defense model as follows.

- *Two kinds of actors are trying to enter the computer* There is some uncertainty about who is who. The legitimates are the customers, and the others are the attackers who are trying to enter the core to do some damage to the company's operation, or to extract information to their benefit. The defenders have some information about the attackers, from which they can derive a probability that the attackers are attempting to enter the system.

- *The defenders' job is to look for and identify the actors* They do so by detecting their "kill chain"; that is, the sequence of attack steps that attackers can implement.

- *Attack steps* The attackers' behaviors are represented by the attack steps. There are uncertainties about the nature of the "gates" (are they really part of an attack?), but there exists some information from previous observations of the attackers' kill chain and behaviors.

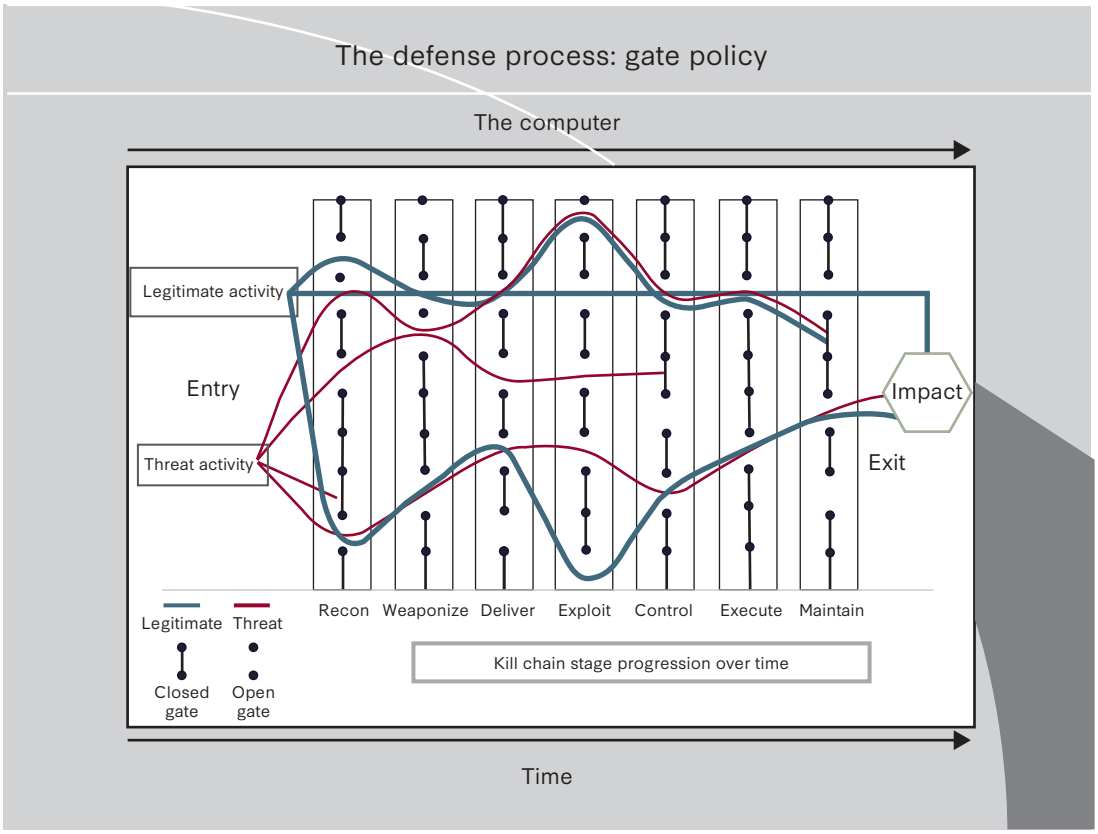
- *The objective of the defenders* The objective is to develop a rational gate policy to optimize the system's value as a decision rule.

- *The defenders' goal* The goal is to lock out the attackers and let the customers pass; therefore, to observe, recognize, and respond to the attackers' "kill chain," stop them but make sure that the buyers can pass. This involves managing the trade-off between the risk of an attack and possible loss of a sale to a customer.

THE DEFENSE PROCESS

The defense is based on a gate policy as described in figure 13, which represents the dynamics of actors' and defenders' moves through a computer. On the left side, actors enter the computer. Some are attackers (thinner path line in the figure) and some are legitimate

FIGURE 13 A defenders' gate policy for cybersecurity



Source: Faber and Paté-Cornell 2020

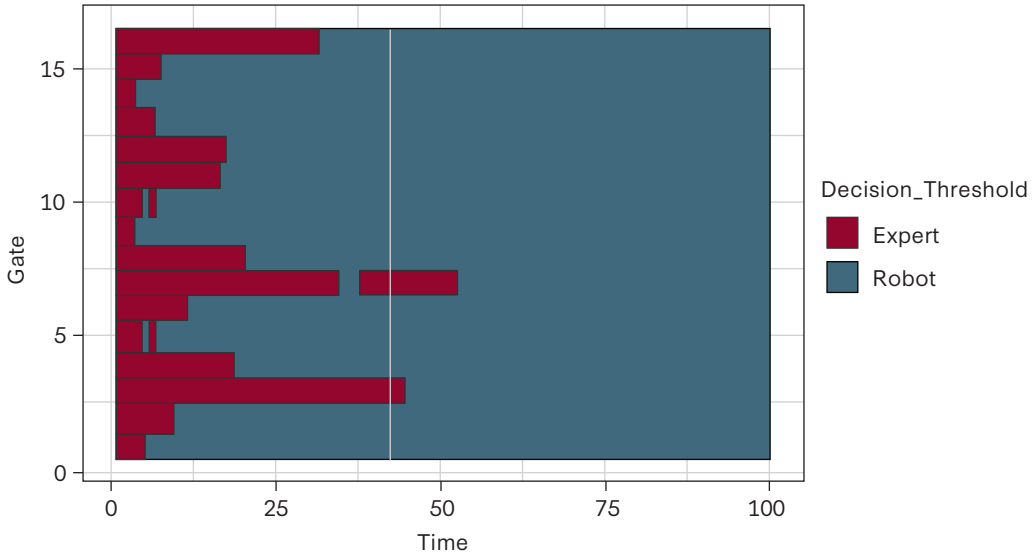
customers (thicker path line). Their goal is to move through the computer, from left to right through the computer representation of the figure to reach its core. The sequence of the kill chain elements is the following: recognition (of a target), weaponization (development of an attack), delivery (of the attack software), exploitation, control, execution, and maintenance. The defenders know these kill chain steps and want to recognize them so that they close the gate when they detect the attempted passage of an attacker. But uncertainties remain in the detection of the kill chain and the nature of the actors trying to pass.

When the attackers implement each of the attack steps, the defenders may receive signals that indicate the danger, but with uncertainty. Based on what they know, they make a rational decision under uncertainty to close the gate or leave it open. In the analysis of their decision, probabilities represent the information available to the defenders. In the example of figure 13, the attackers are stopped three times but make it once to the core. The customers are stopped once but reach the core twice.

HOW THE ROBOT LEARNS

In the example presented in figure 13, the human and the robot make defense decisions to protect the computer at each gate. Figure 14 shows a purely illustrative example of which one

FIGURE 14 The respective roles of the human and the robot in defense of the gates over time. An illustrative example



Source: Faber and Paté-Cornell 2020

of the defenders (the AI system or the human) makes the decision to open or close the gates. Sixteen gates (on the vertical axis) are considered over 100 time units. The decisions of the human are represented in red, and of the robot in blue. In the beginning, the human expert has the knowledge and experience, and the robot learns by observing how the human makes the system's protection decisions.

As shown in figure 14, the robot learns quickly about gates 1, 5, 9, and 14, and takes over the decisions related to these gates after that. It takes more time to learn enough to take over the other gates. Gate 7 presents the particular case where the robot, having taken its control at time 27, passes the hand back to the human expert at time 28. That happens because at time 27, the uncertainties or the outcomes exceed the encoded thresholds—i.e., signals that the robot does not know enough to make the gate decisions. After the human intervenes, the robot has learned enough to take back control.

The learning algorithm is thus the following:

- The expert makes initial gate control decisions.
- The robot learns and takes over at different times for each gate.
- When the uncertainties or the outcomes are too great, the robot passes the hand to the expert.
- After that, the robot knows enough to do the risk management.

LESSON FROM THAT EXAMPLE

The lesson in this example is that the risk of a cyberattack can be managed by an AI system with human supervision. In this example, it was with a human in the loop for the case where the AI framework is not yet equipped to handle a particular situation. The principle is that elements of the kill chain are observed, although with uncertainties, by the defenders who make rational gate decisions given what they know when they observe a signal.

SOME OF THE RISK ANALYSIS CHALLENGES

From the examples presented earlier, a number of lessons can be learned about what to do and what not to do in a risk analysis.

BASIC TASKS AND CHALLENGES

- *The basic task is the formulation of the model* It involves integrating hardware and software, as well as human and management factors. When designing an AI-based risk management model, the challenge is to make sure that the limits of the analytical framework are right and involve the information that is needed. This implies not cutting wrongly for convenience the considered information, including the options that can be envisioned. What matters is the data that one needs, not necessarily the data that one has.

- *Gathering and processing the information* It can also be a challenge. It involves identifying the relevant data, statistics if they exist, and models of the risk and risk management options, including the choice of experts. Note that it is important, if feasible, for the risk analyst to spend time at the location of the operations. As an example, information about anesthesia was gathered in good part by the researchers spending several days in an operating room to observe how the system worked, what could go wrong, what signals would appear, how they could be observed and communicated, and what the response of the different actors might be.

- *Communicating the risk analysis results* That often involves explaining probability and Bayesian reasoning, and especially events dependencies. The basic formula of Bayesian logic is that the probability of [A and B] is that of A multiplied by that of [B given A], and that logic often has to be explained. Both joint and conditional probabilities are elements of the assessment of the probabilities of scenarios, but often confused, including by professionals.

- *Understanding the risk attitude of the decision maker* It is essential to understand risk tolerance as part of the individual's values.

- *Probability goes against deep-seated, deterministic ways of thinking* Few people are comfortable with uncertainties in the first place, and even more so with probabilities and combinations of probabilities. The concept often requires communication and education.

SOME CLASSIC ERRORS IN RISK ANALYSIS

It is useful when doing a risk analysis to be aware of and to avoid some classic errors.

- *Making the problem more complex than needed* It is a common issue since the description of each scenario could be enlarged by adding details that may not be critical to the risk. It is often tempting to add these details because they seem important to the decision maker, but the analysis may show that they may have little impact on the risk result.

- *Mixing uncertainties (facts) and preferences* Guesstimates generally involve some facts. But the probabilities are often biased by what is preferred or feared by the decision maker, with the thought that probabilistic estimates, biased by his or her preferences, will influence the decision toward the choice that they prefer.

- *Irrelevant statistics* They are often used because they are there but have to be avoided if they are no longer representative of the situation—for example, because they are based on values (e.g., financial) of the past and things have changed since then.

- *Assumption that things are designed, constructed, and operated as they are supposed to be* Many failures occur because that is precisely not the case. Such a common misunderstanding of the system may yield risk results that are inferior to reality because they do not account for human errors that may affect the system or the operations in several stages of design and implementation.

- *Irrelevant assumptions of independence* The factors of scenarios are often assumed (without checking) to be independent. Therefore, the probability of their conjunction, if it is estimated as the product of their marginal probabilities, is undervalued. That is because, as mentioned earlier, the probability of [A and B] is that of A multiplied by that of [B given A], which might be greater than the product of the marginals p[A] × p[B] if B and A are highly correlated.

- *Manipulation of the results of a risk analysis to influence a decision* For example, providing a "conservative" estimate out of "prudence" is often tempting to analysts or experts, who are trying to influence a risk management decision by providing manipulated risk results. Yet, it is false and simply leads to wrong risk results and decisions.

FINAL CONSIDERATIONS

In managing risks on the basis of quantitative risk assessments, one needs to consider carefully the impacts of one's decisions.

- By providing risk analysis results, one creates a culture in which the risk of failure is not only recognized but quantified. First, one must recognize that the risk analysis may not be error-free. Yet the quantitative method provides an estimate, and the result may be

small. But in any case, the process and the results may be different from perceptions influenced by a culture of fear, in which the risk may be overstated, or from an environment of denial, in which the risk is understated for a number of reasons, either personal or managerial. The main reason for the quantification is that the objective is not only to state the existence of a risk but to manage it as well and as economically as possible.

- The results of a risk analysis may reveal some effects of the organizational structure. Are there organizational silos that act against others? Part of the risk management may be to modify the organization and/or the process—for example, to make sure that the risks are managed by people close to the problem. In some cases, the risk needs to be brought to the attention of the organizational top, with an accurate definition of the level of severity.

- Finally, a key factor is the quality of the communication system: What information can the risk analysts provide, especially to people who are not familiar with the quantification method and with probability and yet will benefit from a numerical result? Will it be understood when it matters?

In conclusion, when the system is simple enough, risk management may be automatic, and the system is unlikely to fail. In most cases though, perceptions of the failure risk may be somehow inaccurate, and the quantification, if it is well done, provides a solid improvement to simple guesstimates. So, it is important to reward those who do it well!

APPENDIX

CONSTRUCTION OF AN INFLUENCE DIAGRAM AND AN EXAMPLE OF SEISMIC PROTECTION

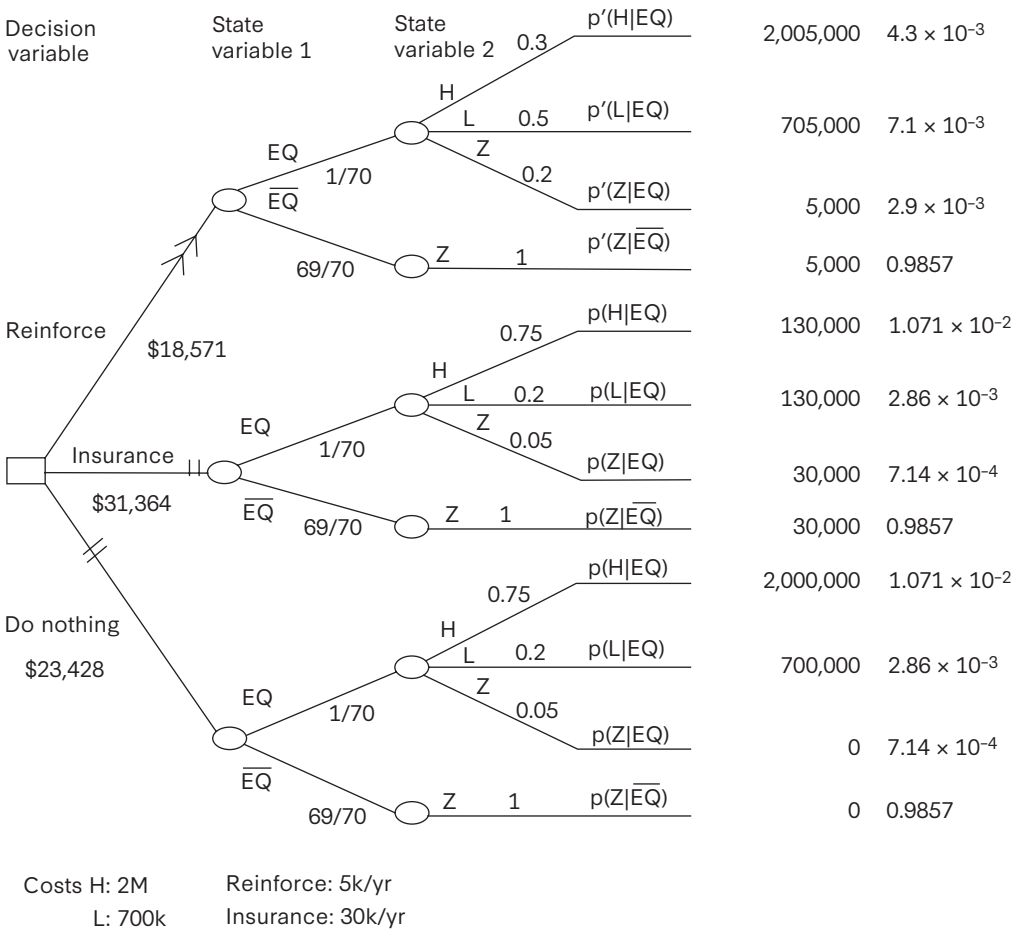
One way to describe the process of constructing an influence diagram is to relate it to the equivalent decision tree for a decision under uncertainties.

1. Decision Trees

Consider a decision that involves several possible options, each leading to an uncertain outcome, and the goal of the decision maker who wants to choose the option that maximizes the expected utility.

The structure of a decision tree is represented below, along a horizontal axis, for the specific example of either reinforcing or insuring a house against earthquakes.

FIGURE A1 The decision tree for the example of seismic risk management focusing on a house



The tree starts on the left with a square decision node that represents the possible options (here: insuring the house, reinforcing it, or doing nothing). Following each option, there are possible sequences of certain or uncertain events (scenarios).

Each event or factor is represented by an oval node linked in a chain to other events in the scenario that are connected to it by conditional probabilities. Each of these nodes represents an event or state variable with its possible realizations (e.g., it happens or not) and the probabilities of each realization conditional on the previous events in the chain.

At the end of the chain, the outcome is represented by its value in the relevant measure (e.g., a monetary value). As mentioned earlier, the decision tree shown here represents the decision to reinforce or insure a house given the possibility of an earthquake (EQ), and the (uncertain) consequences of that earthquake in terms of damage to the house (high H, low L, or zero Z). The vertical line in the probability domain represents the chances of the event on the left, "given" the event on the right.

The probabilities p and p' represent the chances of different loss levels given the decision represented at the beginning (reinforcement or not) and the events that follow (earthquake or not). The numbers on each branch represent the numerical values of the probabilities, and at the end of each path, the value of the cost incurred in the scenario, with or without an earthquake, and the overall probability of that scenario (the product of the conditional probabilities on each branch of it).

- If the decision maker is an expected-value decision maker, he/she chooses the reinforcement solution that minimizes the expected costs.

- If the decision maker is as risk-prone as possible, he/she maximizes the maximum benefits and does nothing.

- If the decision maker is as risk-averse as possible, he/she minimizes the maximum losses and chooses the insurance.

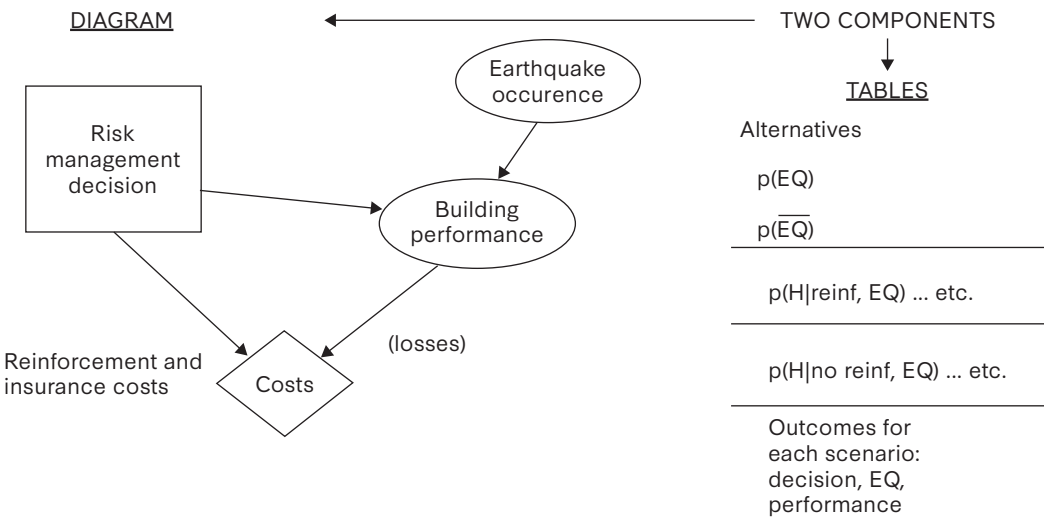
2. Influence Diagram

The corresponding influence diagram can be represented as follows:

The influence diagram of figure A2 has two components: the graph that represents the variables (in squares, ovals, and trapezoids) and their dependencies (the arrow), and the tables that represent the numerical values of the probabilities and the outcomes.

The graph is constructed in the following way. First, the decision is represented in the square or rectangle (here, the risk management options). Then the different events are represented

FIGURE A2 Influence diagram homomorphic to the decision tree of figure A1



in ovals. In this example, the earthquake occurrence does not depend on other variables (no arrow into that node) but it influences the building's performance (the level of damage, H, L, or Z). Both the initial decision's cost and the cost of the damage, if any, influence the overall costs outcome.

The tables contain the values of the probabilities of the earthquake in a given time frame, of the levels of damage conditional on the earthquake, and the chosen risk management option.

This pattern of diagram construction (decision node, variables, and outcomes) can be applied to any decision under uncertainties, and yields the optimal decision given the decision rule introduced in the data; for instance, minimize the overall expected value (or another function) of the costs.

NOTES

1. The risk is sometimes understood to include the possibility of positive as well as negative outcomes of an event. The focus here is on losses, but the method presented below can include some benefits as well.
2. This report is written without equations and very few numbers. A risk analysis, however, is a quantitative exercise and the appendix is analytical and quantitative, presented to show the design and the use of influence diagrams.
3. "Anesthetist" here could refer either to a nurse-anesthetist or an anesthesiologist, who is a medical doctor.
4. The risk attitude is the ratio of the second to the first derivative (-U"(x)/U'(x)) of the utility function U(x) of the spectrum of possible outcomes x.
5. This case has been developed by Col. Isaac Faber for his PhD thesis in risk analysis at Stanford under the supervision of the author (Faber and Paté-Cornell 2020). This section is based on their work.

REFERENCES

References Cited

Abbas A, Howard R. 2015. Foundations of decision analysis. Pearson.

Apostolakis G. 1990. The concept of probability in safety assessments of technological systems. Science. 250(4986):1359–1364. doi:10.1126/science.2255906

Aven T. 2008. Risk analysis: assessing uncertainties beyond expected values and probabilities. Chichester; Wiley.

Bier VM, Lin SW. 2013. On the treatment of uncertainty and variability in making decisions about risk. Risk Analysis. 33(10):1899–1907. doi:10.1111/risa.12071

Bunn M. 2006. A mathematical model of the risk of nuclear terrorism. The Annals of the American Academy of Political and Social Science. 607(1):103–120.

Cooke RM, Shrader-Frechette K. 1991. Experts in uncertainty: opinion and subjective probability in science. Oxford University Press.

Coombs CH, Pruitt DG. 1960. Components of risk in decision making: probability and variance preferences. Journal of Experimental Psychology. 60(5):265–277. doi:10.1037/h0041444

Cornell CA. 1968. Engineering seismic risk analysis. Bulletin of the Seismological Society of America. 58(5):1583–1606. doi:10.1785/BSSA0580051583

Cornell CA. 1980. Some thoughts on systems and structural reliability. Nuclear Engineering and Design. 60(1):115–116. doi:10.1016/0029-5493(80)90263-0

de Finetti BD. 1974. Theory of probability: a critical introductory treatment. Wiley.

Ding Y et al. 2020. Probabilistic assessment of the failure risk of the Europa Clipper spacecraft due to radiations. Risk Analysis. 40(4):842–857. doi:10.1111/risa.13439

Epstein W. 2011. Tsunami hazard and risk assessment [accessed 2025 May 19]. https://woody.com/blog/2011/11/07/tsunami-hazard-and-risk-assessment/

Faber I. 2019. Cyber risk management: AI-generated warnings of threats [dissertation]. Stanford University.

Faber I, Paté-Cornell ME. 2020. Warning and management of cyber threats by a hybrid AI system (robot and operator). Proceedings of PSAM15.

Garrick BJ. 2008. Quantifying and controlling catastrophic risks. Academic Press.

Hora SC. 2007. Eliciting probabilities from experts. In: Advances in decision analysis: from foundations to applications. Cambridge University Press; p 129–153.

Kahneman D, Slovic P, Tversky A. 1982. Judgment under uncertainty: heuristics and biases. Cambridge University Press.

Kaplan S, Garrick BJ. 1981. On the quantitative definition of risk. Risk Analysis. 1(1):11–27. doi:10.1111/j.1539-6924.1981.tb01350.x

Kucik P, Paté-Cornell ME. 2012. Counterinsurgency: a utility-based analysis of different strategies. Military Operations Research. 17(4):5–23.

Lerner JS, Keltner D. 2001. Fear, anger, and risk. Journal of Personality and Social Psychology. 81(1):146–159. doi:10.1037/0022-3514.81.1.146

Lichtenstein S, Slovic P, editors. 2006. The construction of preference. Cambridge University Press.

Merrick J, Parnell GS. 2011. A comparative analysis of PRA and intelligent adversary methods for counterterrorism risk management. Risk Analysis. 31(9):1488–1510. doi:10.1111/j.1539-6924.2011.01590.x

Murphy DM, Paté-Cornell ME. 1996. The SAM framework: modeling the effects of management factors on human behavior in risk analysis. Risk Analysis. 16(4):501–515. doi:10.1111/j.1539-6924.1996.tb01096.x

Ostendorff W, Paté-Cornell ME. 2023. Risk analysis methods for nuclear war and nuclear terrorism. National Academies Press.

Paté-Cornell ME. 1990. Organizational aspects of engineering system safety: the case of offshore platforms. Science. 250(4985):1210–1217. doi:10.1126/science.250.4985.1210

Paté-Cornell ME. 1993. Risk analysis and risk management for offshore platforms: lessons from the Piper Alpha accident. Journal of Offshore Mechanics and Arctic Engineering. 115(3):179–190. doi:10.1115/1.2920110

Paté-Cornell ME. 1999. Medical application of engineering risk analysis and anesthesia patient risk illustration. American Journal of Therapeutics. 6(5):245–255. doi:10.1097/00045391-199909000-00004

Paté-Cornell ME. 2004. On signals, response, and risk mitigation. In: Accident precursor analysis and management: reducing technological risk through diligence. Proceedings of the National Academy of Engineering Workshop on Precursors, National Academies of Sciences, Engineering, and Medicine Press; p 45–59. https://doi.org/10.17226/11061

Paté-Cornell ME. 2007a. The engineering risk-analysis method and some applications. In: von Winterfeldt D, Miles Jr. RF, Edwards W, editors. Advances in decision analysis: from foundations to applications. Cambridge University Press; p 302–324.

Paté-Cornell ME. 2007b. Probabilistic risk analysis versus decision analysis: similarities, differences and illustrations. In: Abdellaoui M, Munier B, Machina MJ, Luce RD, editors. Uncertainty and risk: mental, formal, experimental representations. Springer; p 223–242.

Paté-Cornell ME. 2009. An introduction to probabilistic risk analysis for engineered systems. In: Cochran JJ, editor. The Wiley encyclopedia of operations research and management science. Wiley.

Paté-Cornell ME. 2012. On "black swans" and "perfect storms": risk analysis and management when statistics are not enough. Risk Analysis. 32(11):1823–1833. doi:10.1111/j.1539-6924.2011.01787.x

Paté-Cornell ME. 2022. Warnings and signals: a systems risk analysis perspective. In: Anticipating rare events of major significance. National Academies of Sciences, Engineering, and Medicine Press.

Paté-Cornell ME. 2023. Two different aspects of biases in the use of AI in risk analysis: information versus decision. Proceedings of PSAM 2023 Conference (Probabilistic Safety Analysis and Management).

Paté-Cornell ME. 2025. Alignment of AI systems' risk attitudes, and four real-life examples. The Bridge. 55(1).

Paté-Cornell ME, Dillon RL. 2006. The respective roles of risk and decision analyses in decision support. Decision Analysis. 3(4):220–232. doi:10.1287/deca.1060.0077

Paté-Cornell ME, Fischbeck PS. 1993. Probabilistic risk analysis and risk-based priority scale for the tiles of the space shuttle. Reliability Engineering & System Safety. 40(3):221–238. doi:10.1016/0951-8320(93)90062-4

Paté-Cornell ME, Guikema S. 2002. Probabilistic modeling of terrorist threats: a systems analysis approach to setting priorities among countermeasures. Military Operations Research. 7(4):5–23.

Paté-Cornell ME, Kuypers MA. 2023. A probabilistic analysis of cyber risks. IEEE Transactions on Engineering Management. 70(1):3–13. doi:10.1109/TEM.2020.3028526

Paté-Cornell ME, Kuypers MA, Smith M, Keller P. 2018. Cyber risk management for critical infrastructure: a risk analysis model and three case studies. Risk Analysis. 38(2):226–241. doi:10.1111/risa.12844

Paté-Cornell ME, Lakats LM, Murphy DM, Gaba DM. 1997. Anesthesia patient risk: a quantitative approach to organizational factors and risk management options. Risk Analysis. 17(4):511–523. doi:10.1111/j.1539-6924.1997.tb00892.x

Shachter RD. 1988. Probabilistic Inference and Influence Diagrams. Operations Research. 36(4):589–604.

US BP Commission (United States Commission on the BP Deepwater Horizon Oil Spill). 2011. Deep water: the gulf oil disaster and the future of offshore drilling. Report to the President. BP Oil Spill Commission.

US EPA (United States Environmental Protection Agency). 2005. Guidelines for carcinogen risk assessment. Risk Assessment Forum. Report No.: EPA/630/P-03/001F.

US NRC (United States Nuclear Regulatory Commission). 1975. Reactor safety study: an assessment of accident risks in U.S. commercial nuclear power plants. Report No.: WASH—1400-MR.

Additional References

ASIS (American Society for Industrial Security). 2015. Risk Assessment. Report No.: Standard ANSI/ASIS/RIMS RA.1-2015.

Aumann RJJ, Maschler M, Sterns R. 1995. Repeated games with incomplete information. The MIT Press.

Banks DL, Aliaga JMR, Insua DR. 2021. Adversarial risk analysis. Chapman and Hall/CRC.

Blastland M, Freeman ALJ, van der Linden S, Marteau TM, Spiegelhalter D. 2020. Five rules for evidence communication. Nature. 587(7834):362–364. doi:10.1038/d41586-020-03189-1

Colson AR, Cooke RM. 2017. Cross validation for the classical model of structured expert judgment. Reliability Engineering & System Safety. 163:109–120. doi:10.1016/j.ress.2017.02.003

Cornell CA, Newmark NM. 1978. Seismic reliability of nuclear power plants. In: Probabilistic analysis of nuclear reactor safety. American Nuclear Society.

Daniels M, Paté-Cornell ME. 2014. Quantitative analysis of satellite architecture choices: a geosynchronous imaging satellite example. In: Proceedings of the Space Symposium. Colorado Springs, Colorado.

Dillon RL, Paté-Cornell ME. 2001. APRAM: an advanced programmatic risk analysis method. International Journal of Technology, Policy and Management. 1(1):47–65. doi:10.1504/IJTPM.2001.001744

Dresher M. 1951. Theory and applications of games of strategy. RAND Corporation.

Edwards W. 1953. Probability-preferences in gambling. American Journal of Psychology. 66(3):349–364.

Ezell BC, Bennett SP, Von Winterfeldt D, Sokolowski J, Collins AJ. 2010. Probabilistic risk analysis and terrorism risk. Risk Analysis. 30(4):575–589. doi:10.1111/j.1539-6924.2010.01401.x

Fischhoff B, Slovic P, Lichtenstein S. 1978. Fault trees: sensitivity of estimated failure probabilities to problem representation. Journal of Experimental Psychology: Human Perception and Performance. 4(2):330–344. doi:10.1037/0096-1523.4.2.330

Garber R, Paté-Cornell ME. 2004. Modeling the effects of dispersion of design teams on system failure risk. Journal of Spacecraft and Rockets. 41(1):60–68. doi:10.2514/1.9208

Garber R, Paté-Cornell ME. 2012. Shortcuts in complex engineering systems: a principal-agent approach to risk management. Risk Analysis. 32(5):836–854. doi:10.1111/j.1539-6924.2011.01736.x

Hastie R, Dawes RM. 2010. Rational choice in an uncertain world: the psychology of judgment and decision making. SAGE Publications, Inc.

Haywood OG. 1954. Military decision and game theory. OR. 2(4):365–385. doi:10.1287/opre.2.4.365

Johnson BB, Slovic P. 1995. Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. Risk Analysis. 15(4):485–494. doi:10.1111/j.1539-6924.1995.tb00341.x

Kadane JB, Wolfson LJ. 1998. Experiences in elicitation. The Statistician. 47(1):3–19.

Keeney RL, Raiffa H. 1993. Decisions with multiple objectives: preferences and value trade-offs. Cambridge University Press.

Kucik P, Paté-Cornell ME. 2012. Counterinsurgency: a utility-based analysis of different strategies. Military Operations Research. 17(4):5–23.

Lakats L, Paté-Cornell ME. 2004. Organizational warnings and system safety: a probabilistic analysis. IEEE Transactions on Engineering Management. 51(2):183–196.

Loewenstein GF, Weber EU, Hsee CK, Welch N. 2001. Risk as feelings. Psychological Bulletin. 127(2):267–286. doi:10.1037/0033-2909.127.2.267

McNeil BJ, Pauker SG, Sox HC, Tversky A. 1982. On the elicitation of preferences for alternative therapies. New England Journal of Medicine. 306(21):1259–1262. doi:10.1056/NEJM198205273062103

Morgan MG, Fischhoff B, Bostrom A, Atman CJ. 2002. Risk communication: a mental models approach. Cambridge University Press.

O'Neill B. 1994. Game theory models of peace and war. In: Aumann R, Hart S, editors. Handbook of game theory with economic applications. Vol. 2. Elsevier; p 995–1053.

Paté-Cornell ME. 1986. Warning systems in risk management. Risk Analysis. 6(2):223–234. doi:10.1111/j.1539-6924.1986.tb00210.x

Paté-Cornell ME. 2002. Finding and fixing systems weaknesses: probabilistic methods and applications of engineering risk analysis. Risk Analysis. 22(2):319–334. doi:10.1111/0272-4332.00025

Paté-Cornell ME. 2009. Accident precursors. In: Cochran JJ, editor. The Wiley encyclopedia of operations research and management science. Wiley.

Paté-Cornell ME. 2009. Probabilistic risk assessment. In: Cochran JJ, editor. The Wiley encyclopedia of operations research and management science. Wiley.

Paté-Cornell ME. 2009. Risks and games: intelligent actors and fallible systems. Proceedings of the CESUN meeting.

Paté-Cornell ME. 2012. Games, risks, and analytics: several illustrative cases involving national security and management situations. Decision Analysis. 9(2):186–203. doi:10.1287/deca.1120.0241

Paté-Cornell ME. 2012. Risk management in game situations: principal-agent and adversarial examples. In: Proceedings of the Third International System Symposium. Delft University of Technology.

Paté-Cornell ME. 2015. Uncertainties, intelligence, and risk management: a few observations and recommendations on measuring and managing risk. Stanford Journal of International Law.

Paté-Cornell ME. 2020. Managing failure risks. The Bridge. 50(4):11–12.

Paté-Cornell ME. 2022. Anticipating rare events of major significance. In: Alper J, Hamilton L, editors. Anticipating Rare Events of Major Significance: Proceedings of a Workshop. National Academies Press.

Paté-Cornell ME. 2024. Preferences in AI algorithms: the need for relevant risk attitudes in automated decisions under uncertainties. Risk Analysis. 44(10):2317–2323. doi:10.1111/risa.14268

Paté-Cornell ME, Dillon RL, Guikema SD. 2004. On the limitations of redundancies in the improvement of system reliability. Risk Analysis. 24(6):1423–1436. doi:10.1111/j.0272-4332.2004.00539.x

Paté-Cornell ME, Rouse WB, Vest CM, editors. 2016. Perspectives on complex global challenges: education, energy, healthcare, security, and resilience. Wiley.

Reinhardt JC, Chen X, Liu W, Manchev P, Paté-Cornell ME. 2016. Asteroid risk assessment: a probabilistic approach. Risk Analysis. 36(2):244–261.

Reinhardt JC, Daniels M, Paté-Cornell ME. 2014. Probabilistic analysis of asteroid impact risk-mitigation programs. In: Proceedings of PSAM 12. Honolulu, HI.

Stern PC, Fineberg HV, editors. 1996. Understanding risk: informing decisions in a democratic society. National Academies Press.

31 30 29 28 27 26 25 7 6 5 4 3 2 1

Preferred citation: Elisabeth Paté-Cornell, "Quantitative Risk Analysis: A Number-Free Introduction to the Method, with Examples Including Decision Support from Artificial Intelligence," US, China, and the World Occasional Paper, Hoover Institution, August 2025.

ABOUT THE AUTHOR

DR. ELISABETH PATÉ-CORNELL

Dr. Elisabeth Paté-Cornell is the Burt and Deedee McMurtry Professor in the School of Engineering at Stanford University and a member of the National Academy of Engineering. She currently cochairs the National Academies (NASEM) Committee on Risk Analysis Methods for Nuclear War and Nuclear Terrorism. Her specialty is engineering risk analysis applied to complex systems.

HOOVER INSTITUTION
US, China, and the World

Program on the US, China, and the World

The Hoover Institution's Program on the US, China, and the World delivers data-driven analysis that informs policymakers, business leaders, and the broader public about China's domestic and foreign policy and the bilateral US-China relationship, and provides actionable solutions to the complex challenges at the nexus of the US-China economic and security competition.

For more information about this Hoover Institution project, visit us online at www.hoover.org/research-teams/program-us-china-world.

Hoover Institution, Stanford University
434 Galvez Mall
Stanford, CA 94305 6003
650-723 1754

Hoover Institution in Washington
1399 New York Avenue NW, Suite 500
Washington, DC 20005
202 760 3200

HOOVER INSTITUTION