Shortcomings of the Bow Tie and Other Safety Tools Based on Linear Causality[1]

Prof. Nancy G. Leveson
Aeronautics and Astronautics Dept.
Massachusetts Institute of Technology

Preface: While I am now a professor, I started my career in industry, and when I later moved to university research and teaching, I continued to work with government and industry on real projects. I became involved in safety engineering early in my career after being asked to help a large aerospace company with a high-tech torpedo project. I learned safety engineering from the engineers at the company as it was at that time and, indeed, mostly today, not taught at research universities. I liked it and continued on in the field, believing that preventing accidents and losses was an important life endeavor. In the past 40 years, I have worked in almost every industry and collaborated with a diverse range of specialists on many accident investigations, including those for Deepwater Horizon, the Space Shuttle Columbia, Texas City, one of the Osprey (V-22) accidents, and other lesser-known losses. I have experience in both successful and failed attempts at preventing accidents.

In that same period, the worlds of both business and engineering have increased enormously in complexity and new technology has become ubiquitous, particularly software and autonomy. The traditional safety engineering techniques were never created to handle them. These safety engineering techniques were very useful when they were created in the 1960s but they have become less relevant over the years. The old tools, based on linear causality, however, are still widely used today.

This paper is not about the new approaches being introduced today but instead is an attempt to explain why most of the old approaches are no longer useful and can, in fact, be unintentionally dangerous today. New approaches in both generating causal information and displaying it in a usable way are needed for our current and future systems.

Introduction

To understand the limitations of models based on linear causality, some basic concepts in safety engineering are first presented. Then the standard tools used in safety engineering are described and examples shown of how they greatly oversimplify the causes of accidents, omitting the most important factors and thus underestimate the level of risk in a system. As a result, they provide little help with the engineering of high-tech systems today. The final section focuses on the Bow Tie diagrams that seem to be increasingly used to visualize the results.

The problems and limitations arise with the underlying linear causality model on which the tools are based so this paper starts there.

Theoretical Accident Causality Models

Basic to safety engineering, as is the case with any engineering, is the need for a model to explain system behavior, in this case how accidents occur. Otherwise, we would be faced with a totally random world, with few tools to assist in building systems that achieve our goals without causing harm in the process.

Safety analysis can be divided into two types:

[1] I would like to thank Andrew McGregor of Auckland, New Zealand, for his tremendous help in reviewing and copy editing this paper. Thanks also to Captain Shem Malmquist, who provided comments and assistance.

1. Accident Analysis: Identifying the cause of a particular loss that has occurred in order to take steps to prevent similar losses in the future.
2. Hazard Analysis: Identifying the potential causes of accidents that have not yet occurred, in order to prevent them or, if that is not possible or feasible, to reduce the losses if they do occur.

Note that both goals involve identifying causes and the only real difference is the information that is available at the time. In fact, the second (hazard analysis) can be thought of as "investigating an accident that has not yet occurred."

To perform either type of safety analysis, one needs to have a model of accident causation, that is, the analysis must be based on assumptions about how accidents occur. Models represent our assumptions about how the world operates. For example, if an underlying assumption is that accidents are caused by operator error, then the analysis will focus on what the operators did to contribute to the loss. Such assumptions about the causes of accidents always underlie engineering for safety, but those doing the analysis may be unaware of any subconscious assumptions they are making.

The Linear Chain of Failure Events Model

General models of causality have been proposed and used. The most common is the oldest. It consists of the assumption that accidents are caused by chains of failure events, each failure being the direct consequence of the one before. For example, someone enters the lane in front of your car, you slam on the brakes but are too late in applying them, and therefore you hit into the car in front of you. Perhaps in addition, someone was following too close behind you and rear ends your car.

Figure 1 shows an example of applying a simple chain of events model for a tank explosion. Note that the chain can have logical "ANDs" and "ORs" in it. In this accident, moisture gets into the tank, which leads to corrosion, which in turn causes weakened metal. The weakened metal along with a particular operating pressure leads to a tank rupture, which causes fragments to be projected. The fragments lead to damaged equipment and/or personnel injury.

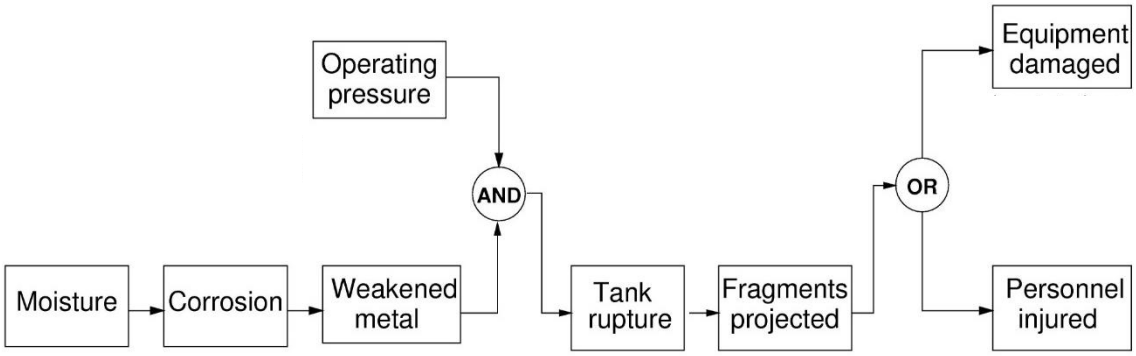


Figure 1: Chain of events model for a tank explosion

Using this model of accident causation, it appears that the simplest way to prevent such an accident is to eliminate one of the events in the chain. An alternative is to put barriers between the events so that the consequences of one event does not lead to the next event in the chain. An example of a barrier in this case is to put a screen around the tank so that in the event of a rupture, the fragments cannot be projected outside a protected area. The use of barriers is common practice in the nuclear power and other industries that use "defense in depth" to prevent accidents. In such approaches, multiple barriers are provided with each barrier provided to backup the previous one. For example, protective cladding is put around the nuclear fuel to contain the radiation. In case the cladding is not

effective, a shutdown system is used to stop the reactor. A final defense, if everything else fails, is the containment structure surrounding the entire reactor building.

Figure 2 shows an annotated model of the same tank explosion accident chain with possible protection or control activities denoted. For example, moisture might be kept out of the tank by using a desiccant or the tank might be coated with stainless steel to prevent corrosion.

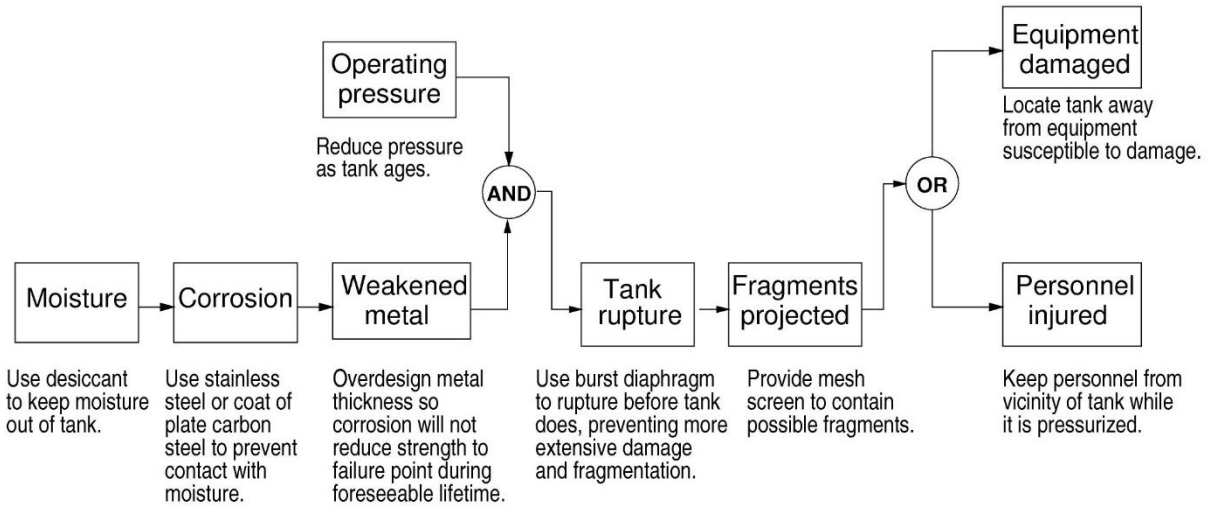


Figure 2: Tank explosion example shown with added protections

There are a few things to note here. First is that direct causality is assumed, that is, each event leads to the next event in the chain. Also, the preceding event is necessary for the following event to occur, i.e., if moisture does not get into the tank, then corrosion will not occur. That is, the previous event in a chain is both necessary and sufficient to produce the following event.

Using this model to explain a loss that has occurred, the analysis works backward from the loss event to identify the causal chain. The initial event is called the "root cause." While the event labeled the root cause does not have to be the first one documented in the chain, it usually is. Note that almost always the stopping point is arbitrary and often is a human operator. In fact, more previous events could be added, which would then be the "root cause." The search works backward until something is found that is easy to prevent or the search cannot easily go backward any farther. That event is labeled the root cause. Sometimes politics and liability become involved in the selection.

As an example of how events could be added, consider the first event in Figure 2, which is moisture entering the tank. That moisture must be introduced somehow and there probably were design features used to prevent water and moisture reaching the tank. The failure of those protection devices could be added to the beginning of the chain. Is that failure the root cause? What is chosen as the root cause is usually somewhat arbitrary as any of the events in the chain could theoretically be labeled as such.

There are no operators in the simple example shown, but usually an operator is selected as the root cause. We might change the example to have an operator opening a valve that allows moisture to get into the tank. One reason that operators are usually chosen as the root cause is that it is difficult to go backward through a person to an event that causes the operator error. The interface design, for example, is not an event. It is a feature of the system design or the context in which the operator is working. What is the direct cause of the pilot giving an incorrect command to the flight management system? This is one reason why operators are usually blamed for accidents, although there are others.

Software also is not included in Figure 1, but we could have, for example, by having software control the burst diagram or relief valve in the tank. While it is easy to say "Software does not open the relief valve" (perhaps in a box following "Tank Rupture," it is more difficult to think of a way to protect against this behavior. Only a very small part of real software can be tested within a reasonable amount of time. Software is an abstraction (set of instructions) that cannot fail—it does exactly what it was told to do so the problem must involve a design or requirements error on the part of the engineers. How does one create simple protections against that?

Note that the other boxes in the chain might also contain design errors (the design of the tank, for example) but those causes are omitted from the chain of events model because they are not events. We will come back to this later. I have noticed that in most real-world hazard analyses, there usually is not much included about software or the design of the product (e.g., the aircraft) or operations as a causal factor. Human error and physical failures are the primary causes considered.

Limitations of the Linear Chain of Events Model

The examples in Figures 1 and 2 are quite simple, of course. Real systems today may have hundreds or even thousands of such chains of events leading to losses. I was told of one fault tree (which generates causal chains, just as do all hazard analysis techniques) for the Integrated Modular Avionics system for an aircraft that required over 2000 pages to document the results. And this was only for one part of the aircraft. During Space Shuttle development, a FMEA (Failure Modes and Effects Analysis) identified 40,000 critical items. It's not clear what to do with the information that the failure of 40,000 individual items could lead to the loss of the Shuttle, but only a government project like the Shuttle could have the resources to identify all of these, let alone provide protection against them. And, of course, Shuttle design errors and poor management decision making are omitted from this analysis. This omission includes the causes attributed to the actual two Space Shuttle losses. The only solution to dealing with complex systems using these failure-event-chain models is to either simplify the models or omit most of the factors contributing to the accidents.

There are other inherent limitations of this traditional and almost universal chain-of-failure-events model of accident causation. First, there is an assumption that the events and barriers fail *independently*, that is, there is nothing that will reduce or eliminate the effectiveness of all of them at the same time. Therefore, it is assumed that the risk of an accident, if all defenses are implemented correctly, is low. However, the independence assumption is untrue. For example, accidents commonly occur because budget cuts, demands for increased productivity, or competitive pressures make all the barriers ineffective at the same time. A poor organizational or safety culture (e.g., management pressures to ignore safety rules and procedures) can also undermine the effectiveness of the safety controls and the applicability of the model. These so-called "systemic factors" do not appear in the models and, for the most part, have to be ignored to perform quantitative or even qualitative risk assessment.

Another critical omission are accidents that involve non-failures, where all the components may operate as designed, but their interactions lead to the failure of the system as a whole. Accidents resulting from the unsafe Interaction of non-failed components may stem from complexity in the overall system design and the incorporation of software controls and autonomy into the design. Here is an example: An A320 landing at Warsaw airport during a rainstorm could not stop and crashed into a small mound at the end of the runway. The flight crew tried to activate the reverse thrusters (i.e., the temporary diversion of an aircraft engine's thrust so that it acts against the forward travel of the aircraft) but the software would not let them do so because the software "thought" that the aircraft was still in the air. It is dangerous to activate reverse thrust when airborne so protection had been built into the software to prevent this from happening, i.e., to protect against a flight crew error or perhaps some

type of strange failure mechanism in the engine leading to reverse thruster activation while the plane was airborne. But the engineers did not fully account for all possible environmental conditions at an airport, even the unusual ones. What failed here? Certainly not the flight crew or the software, both of which did exactly what they were instructed to do. This example highlights a problem that often occurs in complex systems today, namely, it is difficult and often impossible to identify all potential conditions that will occur in operation and all possible behavior of the system itself. Engineers call these the "unknown unknowns." Most of the accidents I see today are a result of these types of design errors, although they are often incorrectly blamed on the pilots or human operators.

The simple chain-of-events model also does not account for changes over time. Nothing is static, not even hardware and software design (which both need to be maintained and changed over time) and certainly not social systems. So, the protections instituted against accidents may lose their effectiveness over time, even protections that are created to reduce the impact of changes. In Figure 1, one protection strategy included is reducing the pressure in the tank over time. But such reductions may be put off because of productivity concerns. In addition, restrictions on keeping personnel away from dangerous equipment may be relaxed to allow maintenance activities to proceed without disrupting operations by requiring equipment to be shut down before working on or near it. Again, productivity motivations may be paramount. While dangerous equipment may be isolated at first, those restrictions or constraints may be reduced over time. For example, at Texas City the ISOM tower that exploded was at first isolated, but pressures grew for expanding office space and trailers and the only available space was next to the ISOM tower. Formulas for chemicals and the physical composition of equipment may change over time. Humans start to behave differently as they become more familiar with equipment and their job. Shortcuts may start to be taken. Safety culture can even change over time, thus affecting overall behavior. None of this is included in the chain of failure events and the provisions originally provided to prevent the events become less effective. Even if potential changes could be included, it is usually difficult to predict what changes will occur in the future. Aircraft and other systems may continue to be used for decades.

Finally, a lot has been learned about human factors in the last 50 years, with large advances in cognitive psychology, which was only in its infancy 50 years ago. Before we had this knowledge, we could assume that human behavior or "failures" were essentially random and independent of the context in which they occurred. This belief now rarely applies as even simple "slips," such as reaching for one button and accidentally hitting another one that is close by and similarly shaped, cab be reduced or eliminated by simple design of the interface. We know too much today about human cognitive behavior to assume that it is random and independent of contextual factors and design of the system in which the human is working.

Past assumptions also do not fit the role of humans in systems today where the humans are mostly managing complex automation rather than directly controlling physical devices or computer-automated functions. The future will see even more of the human role changing to one of being a manager or monitor of computers and even partnering with automation to achieve common goals as responsibilities are divided between the machine and the human. Autonomy does not usually mean that humans are totally eliminated from systems (except in the simplest cases) but only that their roles are changed. None of these new roles and human factors considerations are included in the simple model of accidents based on chains of failure events. They cannot be represented using a simple linear failure model.

Note that descriptions of chains of events as dominoes falling, holes in swiss cheese, or similarities to men's formal attire such as bow ties, are only graphical differences. The chains of events may be drawn differently, use different notations, or apply different analogies, but they all are describing the same

underlying chain of failure events model. They are not different causal models, but simply different names or notations for the same linear causality model.

Safety engineering has been built on this limited causal mode of how and why accidents occur. Despite these limitations, no alternative to this traditional accident causality model has been suggested until relatively recently.

Analysis Techniques Built on the Chain of Failure Events Model

Creating and using modeling and analysis tools is difficult and costly. To justify the resources needed, the information generated by the modeling and analysis effort needs to be useful for achieving important goals.

While the chain-of-failure-events model of the general accident causal process dates back a long time, the first modern causal analysis tools were created in the 1960s to deal with the complex and potentially very dangerous systems we were starting to build at the time, such as intercontinental ballistic missile systems. A brief summary of these techniques is presented next.

Using Backward vs. Forward Search to Generate Causal Chains

The first problem facing the analyst is how to generate the causal chain of events. It's not at all as easy as it may seem from looking only at some oversimplified examples like the one in Figures 1 and 2. Generating the chain of events after an accident, while fraught with political implications and bias about the cause of an accident, is still easier than generating all possible chains of events needed for proactive hazard analysis and risk assessment activities before an accident occurs. Therefore, proactive generation of causal chains is emphasized here.

Finding the causal chains usually involves some type of search process. Either initial events are followed forward to identify their eventual results or a final bad event is traced backward to its initiating events. See Figure 3 below.

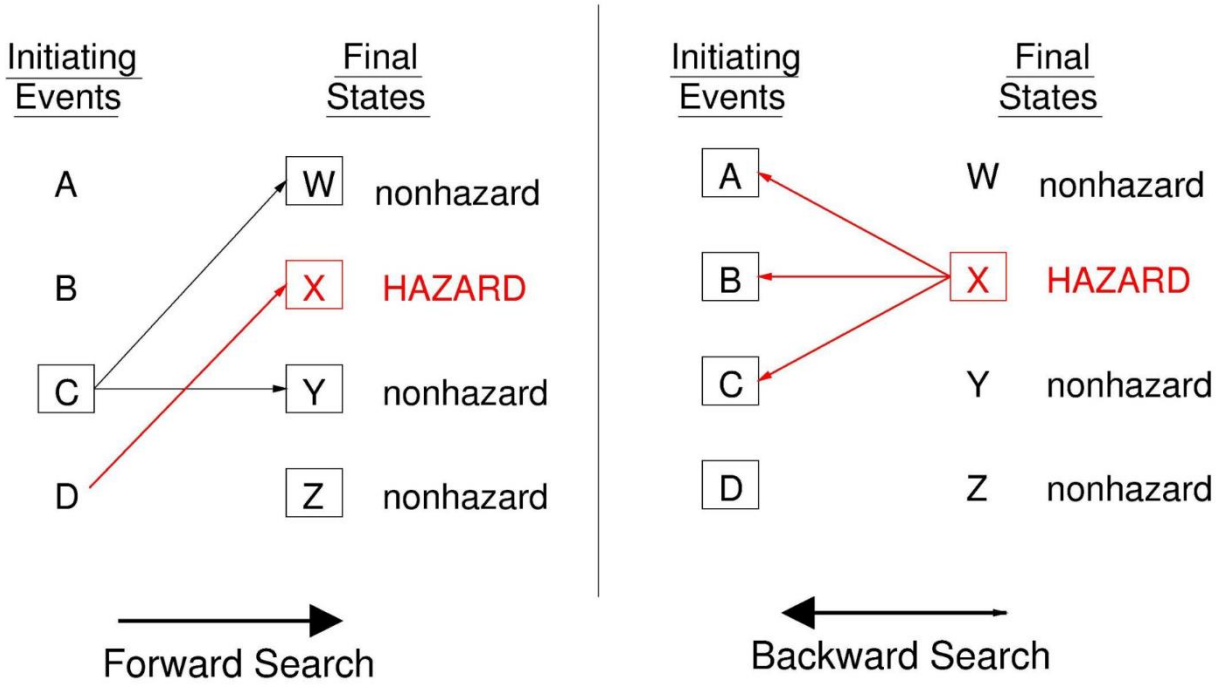


Figure 3: The two types of search used in generating chains of failure event models.

Forward search involves identifying some potential initiating event and propagating it forward to determine whether it can result in a hazard.[2] The problem is that it is inefficient. As seen in the figure, paths to both non-hazardous events will be generated, but only the hazardous ones may be useful in hazard analysis. In addition, the cost of following one event forward may be so great that only single event causes can be considered. All the combinations of all possible initiating events are impractical to use as a starting point.

Backward search, where the search starts with hazardous states and determines how they might be reached, is more practical. In addition, combinations of initiating events may theoretically be identified without extra work. Because of the efficiency of backward search techniques, most search techniques used today are backward. Only relatively simple types or parts of systems are subjected to forward search.

Failure Modes and Effects Analysis (FMEA)

FMEA (or FMECA where the "C" stands for criticality) is a type of forward search technique. Figure 4 is one such example.

Item	Failure Modes	Cause of Failure	Possible Effects	Prob.	Level	Possible actions
Motor Case	Rupture	(1) Poor workmanship (2) Defective materials (3) Damage during trans. (4) Damage during handling (5) Overpressurization	Destruction of missile	0.0006	Critical	Quality control

Figure 4: A FMEA for the rupture of a motor case in a missile.

In this example, the item being analyzed is the motor case of a missile. The initiating event is a rupture of the motor case, and the failure modes of the rupture are identified. Note that the failure modes of the initiating event must be known for FMEA to be performed, and there must be only a limited number for a FMEA to be practical. In the example shown in Figure 4, the effects of a motor case rupture are assessed to be critical, namely the destruction of the missile. Five causes are identified. As this is a forward search technique, all the potential failure modes and effects of those failures must be considered, even those that are not important, in order to find all the important ones. A probability is assigned and potential control or mitigating actions, in this case quality control procedures, are

[2] The term "hazard" is used in this paper as it is used in safety engineering, that is, a state or condition that, under some set of circumstances, will lead to an accident. By definition, the hazardous state is limited to the system under the control of the designers or operators of the system. The goal of safety engineering is to eliminate or mitigate the dangerous effects of a potential hazard. For example, in aviation, a mountain is not a hazard as it cannot, in almost all cases, be eliminated. The related hazard is an aircraft coming too close to the mountain, which the designers and operators of the aircraft can prevent. A similar example is weather. We can do little about changing the weather, but we do have control over whether the aircraft comes in contact with dangerous weather events or, if it does, protect against the impact of the weather on the aircraft.

described to prevent or control the failure. In a more realistic and complete example, a different possible action might be identified for each of the identified causes of the failure.

As one can imagine, the results of such a forward analysis in complex systems (a B737 has 367,000 parts, a B747 has six million parts where half of those are fasteners, and an A380 has about 4 million individual parts) can be enormous and extremely expensive to produce. Usually, only single failures are considered because including all combinations of failures makes the technique infeasible to perform on any but the simplest of systems. Alternatively, abstractions might be used but they must necessarily be very high-level and therefore tend to not be very useful.

Although functions may be considered rather than physical parts, the amount of work is still enormous and can take hundreds or even thousands of pages to document. In addition, the probability of failure of parts may be known but not the probability of functions, which may be new or newly designed for a particular system. We usually create new systems because we want to add or change functionality, not employ the same functionality we had already. Finally, if software is included, the "failure modes" might include any potentially incorrect outputs, timing, or sequence of behavior and thus may be so many that it is not feasible to include them all. Using abstraction to reduce them to a reasonable number means leaving out important information. Using abstractions such as "correct" and "incorrect" output is useless. Therefore, claims of FMEA being useful for software are doubtful.

The output of an FMEA is usually documented in a tabular form because of the enormous amount of data involved. Figure 5 shows a graphical model of the event chains identified for the simple and partial example of Figure 4. Of course, such a graphical notation would be totally impossible to use for more than just a small part of the whole missile system. The type of notation shown in Figure 5 is equivalent to that used in Bow Tie diagrams (described later), which got the label because of its similarity to men's bow ties. For Bow Tie diagrams to be practical, only small parts of the system can be shown or the event chains have to be simplified to the point where they provide an incomplete and misleading view of causality.

When probabilities are known, a quantitative analysis of the failure events (hazards) can be performed, again assuming everything is independent, which is almost never the case.

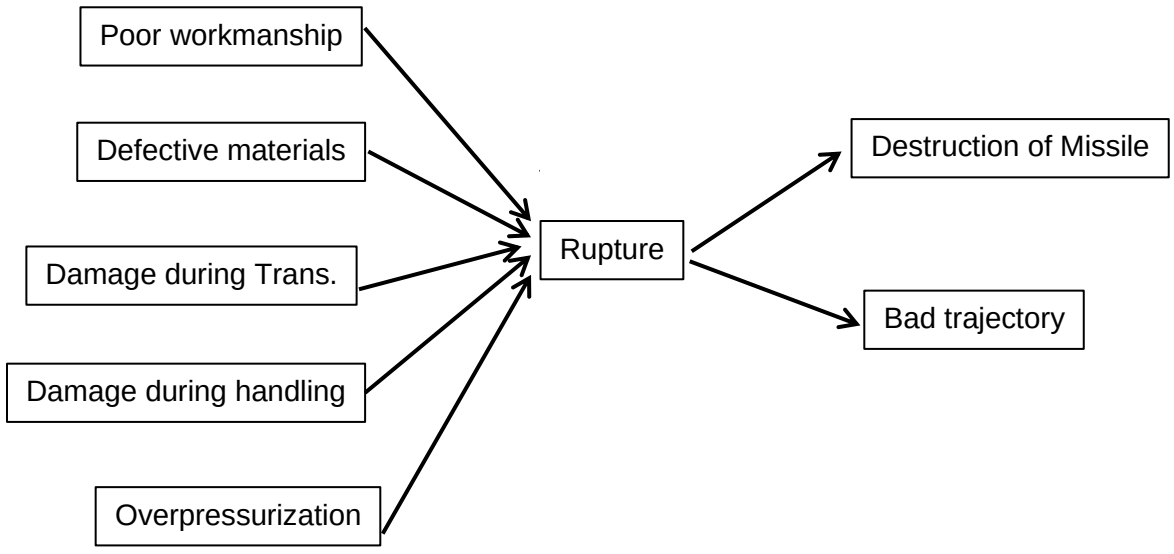


Figure 5: The bow tie diagram for the FMEA results shown in Figure 4 with the controls (possible actions) omitted for clarity.

Remember, FMEA and the other tools described here were created when the systems we were building were very much simpler and used different technology than that used today. Problems in scalability were not as relevant as they are today.

To summarize, in a FMEA, only single failure events are included; it works best on hardware and mechanical system components, not software, human operators, or organizational factors; it can be inefficient as it analyzes the important along with the unimportant; it tends to encourage redundancy as a solution (which may be inefficient, ineffective, and very costly); and the failure modes must already be known so it works best for standard parts with few and well-known failure modes.

Of course, the analysis could be simplified by leaving out details or working on larger components or functions such as failure of the entire motor, but that also limits its usefulness as the details are important in deriving cost-effective controls. And, once again, context, which is critical for evaluating safety, is omitted. The Inertial Reference System in the Ariane 4 spacecraft launcher was safe when used in that system but when used in the Ariane 5 resulted in its destruction. Individual components or functions are not safe or unsafe; safety only is meaningful at the system level. FMEA analysis is therefore most relevant for system reliability analysis, not system safety analysis.

Fault Tree Analysis

Fault Tree Analysis, perhaps the mostly widely used hazard analysis technique, was created in 1961 by Bell Labs and Boeing to analyze the Minuteman missile system. It is a top-down or backward analysis method in which the starting point is the hazard to be avoided and the chains of failure events leading to that hazard are identified. Fault tree analysis is therefore more practical for large or complex systems than a forward analysis technique like FMEA.

The results of fault tree analysis show the chains of events connected in a tree structure but it is no different than individual chains being constructed separately except that the notation saves space by being able to share boxes. A simple example is shown in Figure 6, which has three chains of events specified: The rupture of the hot water tank can result from

1. *Failure the temperature device to actuate the controller* AND *Failure of the relief valve to lift*.
2. *Failure of the controller to actuate the gas valve* AND *Failure of the relief valve to lift.*
3. *Failure of the gas valve to close* AND *Failure of the relief valve to lift.*

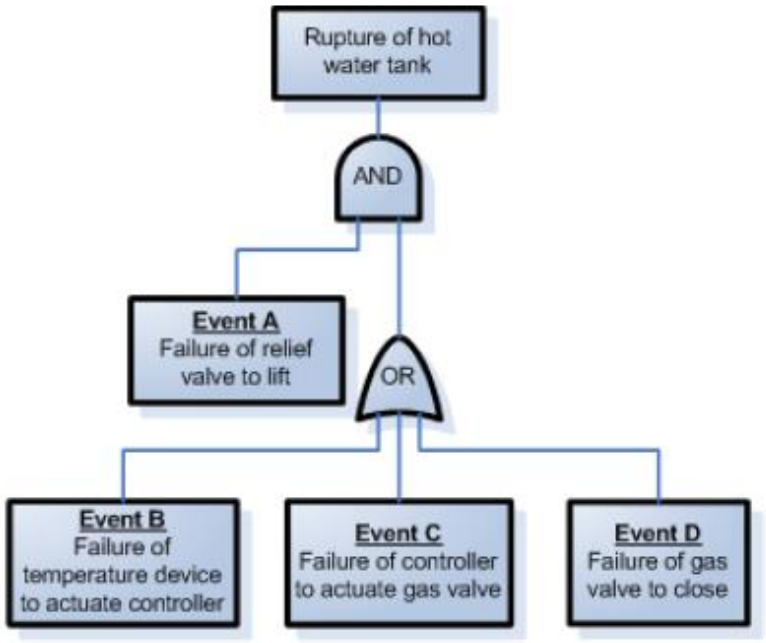


Figure 6: An example fault tree from the original Bell Labs study.

The fault tree shown in Figure 6 is equivalent to Figure 7, which again is similar to the "bow-tie" like notation of Figure 5. Note that only the left side of the bow tie appears because fault trees start with a hazard (not the accident) and work backward.

The fault tree itself shows only the result of the analysis, which is done in the head of the analyst. There is no model of the system on which it is performed nor any procedure to follow. The tree shows the final results. Again, because the tree is constructed of simple logic statements, a probability of the top failure event can be calculated assuming the failure probabilities are known for all of the boxes and the failure events are all independent.

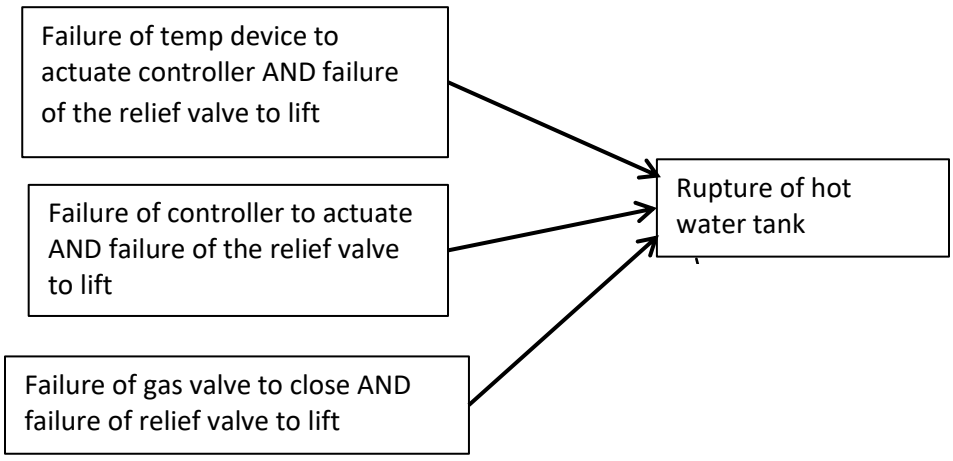


Figure 7: The equivalent bow tie diagram for the fault tree analysis results shown in Figure 6.

Lest the reader get the impression that only extremely simple systems can be analyzed using fault trees, Figure 8 shows a small part of the fault tree created by MITRE in 1983 for the U.S. FAA

certification of TCAS (Traffic alert and Collision Avoidance System). TCAS is used on virtually all commercial aircraft today. Because there was no way to obtain probabilities for most of the boxes and they were certainly not independent, no attempt was made to quantify the resulting model. This fault tree was used to identify the causes of hazards and to design TCAS and its procedures so that it could handle the causes identified. After developing new, more powerful hazard techniques in the past few years, I have discovered that the TCAS fault tree was not complete, which is a problem with fault trees and failure event chain causality models in general. But we did not know that at the time.

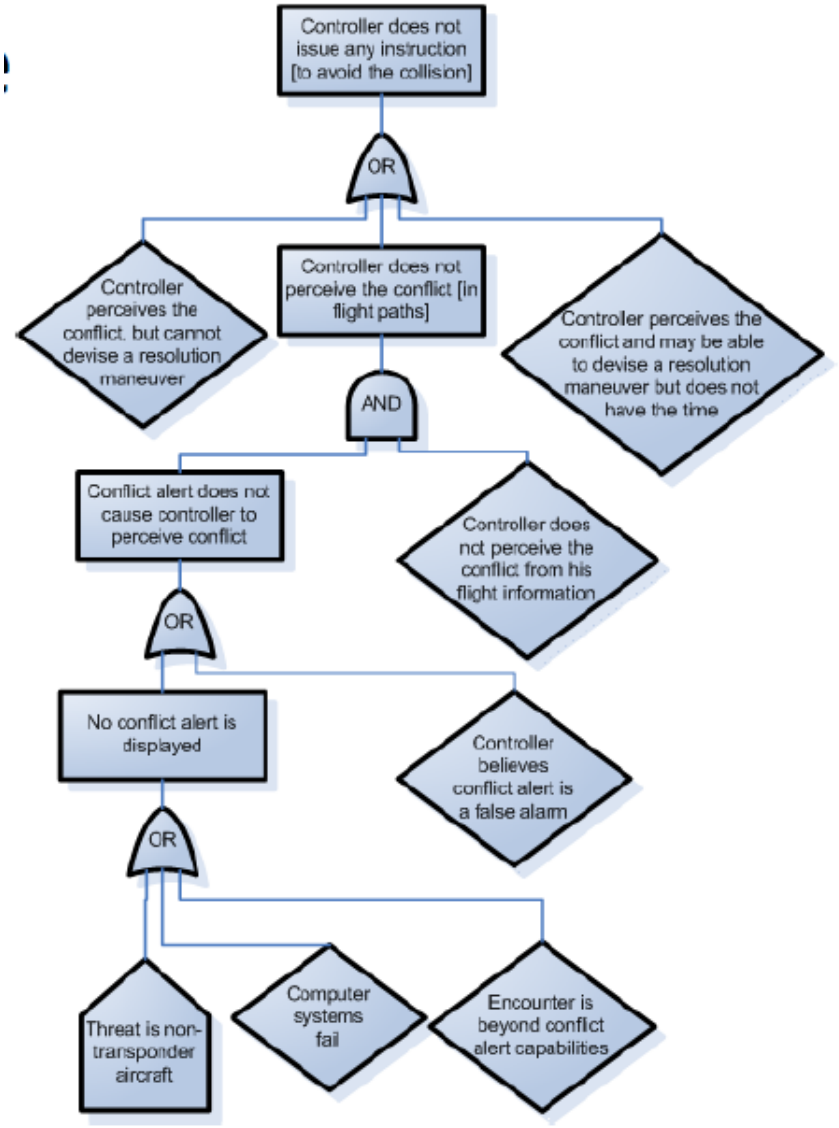


Figure 8: A small part of the actual fault tree created by MITRE in 1983 during the development and certification of TCAS[3]

[3] A reference cannot be provided as I have this only because I was personally involved. The complete fault tree is not publicly available.

There are some good things about fault tree analysis. It captures combinations of failures and is more efficient than FMEA as it is a backward analysis process (starting from the hazard at the top of the tree) so only the hazardous failures are considered. In addition, information about potential common-mode failures can be used to improve system reliability.

However, most of the general limitations of event chain models exist in fault trees. Independence among events is often assumed, it is difficult to capture non-discrete events, it does not easily capture systemic factors, generating the trees can be labor intensive, there is no underlying common model being used by the analysts to represent the design of the system, they can become complex very quickly and thus difficult to review, and, most important, the same non-failure events are omitted as in the other search techniques. It is not appropriate for social or organizational systems or of much use for software.

Event Trees

In 1967, when an attempt to build a fault tree for nuclear power plants became unmanageable, event trees were invented. This is a forward search technique, but the initiating event is the hazard rather than individual component failures. A typical initiating event might be a pipe rupture or overheating of the nuclear core. The event tree then shows all possible outcomes that might result from the hazard and failures of the protection devices. The probabilities are inserted into the event tree. Figure 9 shows the format of an event tree (where the probabilities would appear in the locations where "Succeeds" or "Fails" appear). Figure 10 translates that figure into the bow tie notation. In this case, only the right side of the bow tie is shown as event tree analysis starts with the hazard (called the initiating event in Figure 9) and generates the rest of the chain.[4]

In Figure 9, the first event (the hazard) is a pipe break. The second failure event involves whether electrical power is available as power is required for the operation of the rest of the recovery actions except the final static containment structure. The first layer of defense is the Emergency Core Cooling System, then fission product removal, and finally the integrity of the overall containment vessel.

Only showing the event chain after the hazardous event makes sense in this context as nuclear power plant design, or at least certification for safety, is focused on the recovery of the plant after a hazard occurs and not preventing the hazard from occurring. The use of protection systems (called "Safety Systems") contrasts to the approach in other industries, such as fail-safe design in aviation, where the goal is to prevent the hazard. Aircraft usually do not have a safe state to easily move into from a hazard (such as a loss of propulsion), although recovery actions for some aircraft hazards are possible given highly skilled pilots, the provision of the ability to execute the recovery actions is designed into the aircraft, and often a lot of luck.

Event trees are only practical and useful for systems in which the incorporation of protection systems (moving to a safe state) is possible. Strangely, attempts have been made to use event trees in aviation, particularly air traffic control, which makes little sense as there usually is not a safe state to which to revert, and, in addition, the ordering of failure events is not fixed.

[4] Fault trees and event trees are often combined, but in a different way than as a simple event chain.

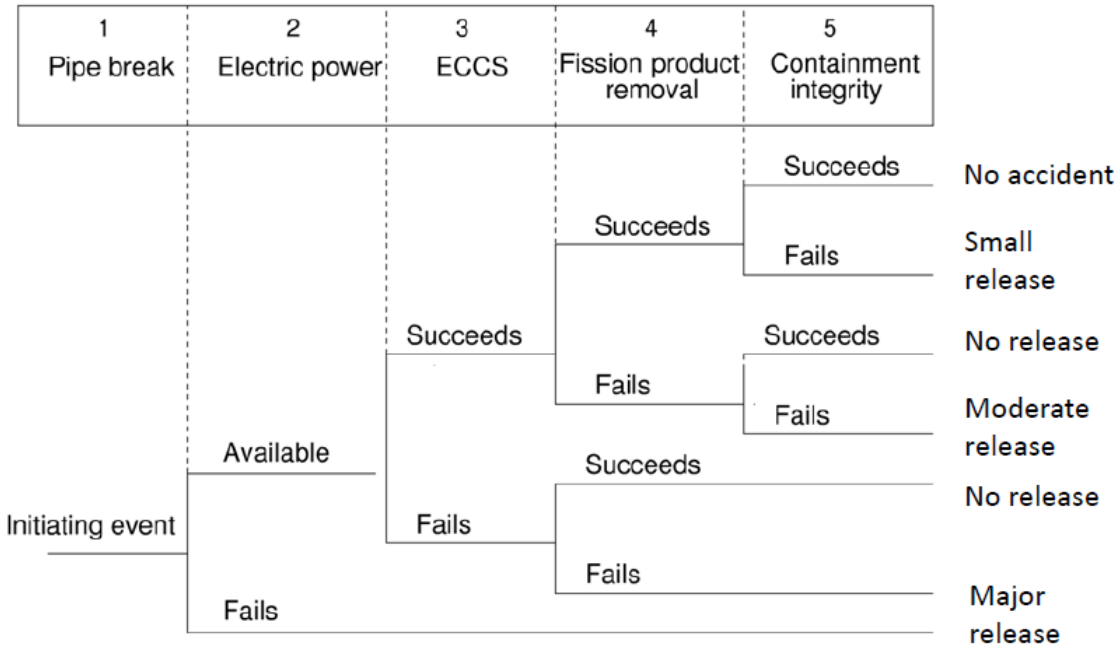


Figure 9: An example event tree for a nuclear power plant.

As can be seen in Figure 10, the event tree is equivalent to the right-hand side of the bow tie diagram. Once again, the events must be independent, but, in this case, there must also be an ordering of potential events over time (i.e., the chronology of events is stable, that is, their sequential order must not change). In the Bow Tie models I have seen, this limitation is avoided by including only single, independent, unordered recovery events and omitting everything else.

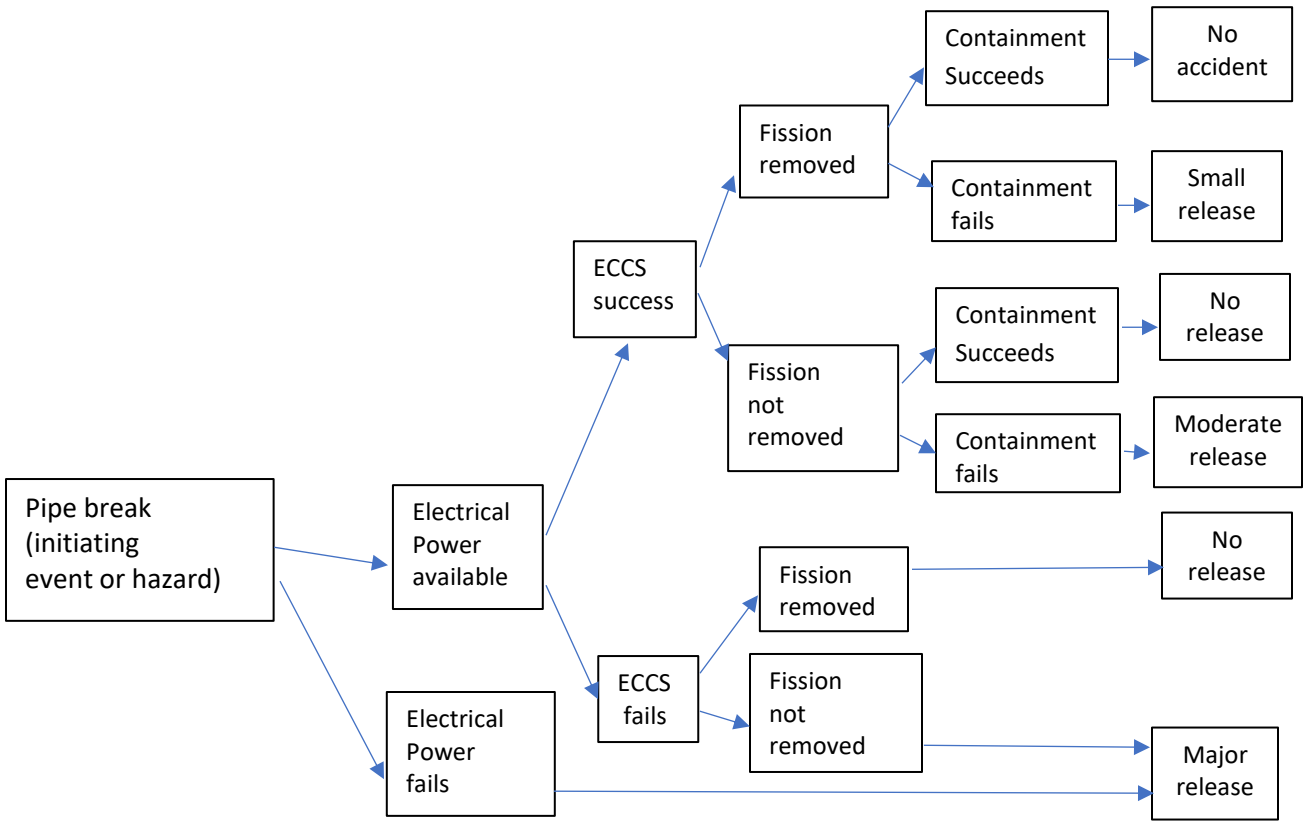


Figure 10: Bow Tie Notation for the Event Tree in Figure 9.

HAZOP (HAZard and Operability Analysis)

In the mid-60's, Trevor Kletz and others in the chemical process industry created a hazard analysis tool called HAZOP that became the most commonly used tool in that industry. This technique is the only one that uses documentation of the system design to generate the failure event chains. For the others, the design is in the head of the analysts. The documentation used is a Piping and Instrumentation Diagram (PID) of the plant so, again, the focus is on physical failures. Attempts have been made to include software and human factors, but none have been very successful. Also, aspects of plant design that fall outside the PID such as building and operator console layouts cannot be easily included in the analysis. And as with all the other techniques, systemic factors in accidents are omitted.

Leaving out the details, HAZOP starts with the analysts hypothesizing different types of deviations in the components of the system documentation, such as a pipe not having enough flow, having too much flow, or having reverse flow. These deviations are expressed as *guidewords* or *prompts* that are methodically posed as the analysis progresses through the PID schematic. The analysts then try to determine what (failure) events could lead to this condition and what events might result from it. The difference here is that the process starts in the middle of the chain of events and works both forward and backward instead of at one end or the other. But where one starts in the chain is irrelevant in terms of generating the entire chains of events, although the ease of generation might be affected. Because chemical plant designs can have lots of physical components and in order to not oversimplify or exclude causes, the results are documented in tables as shown in Figure 14.

HAZOP MINUTES

System No. :			System Name :		Page of
Report of HAZOP Study			Client :	Project :	Content No. :
Date :			Present :		Report by:
SYSTEM NUMBER :			SYSTEM :		
No	Guide Word	Deviation	Possible Causes	Possible Consequences	Action Required

Figure 14: A typical HAZOP form used to document the analysis.

Bow Tie Diagrams

In the early 1970s, several people noticed that fault trees could be combined with event trees to show the entire event chain. The result was first called Cause-Consequence Diagrams by Bob Taylor, who put the fault tree at the top of the page (but upside down) and the event tree below it so the two were connected by the common “hazard” box (called the Critical Event). See Figure 11).

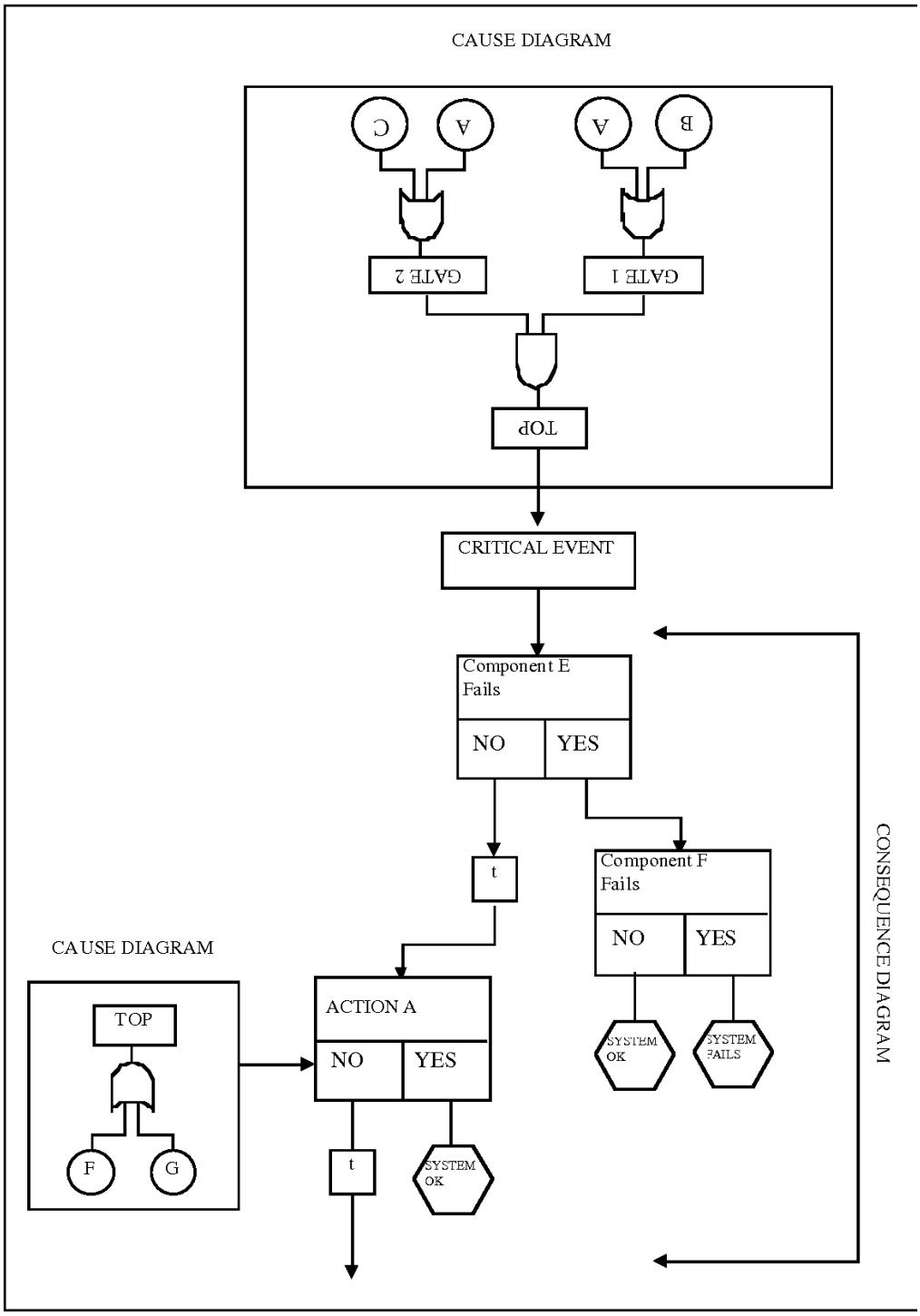


Figure 11: Original Cause-Consequence Diagram with the fault tree at the top and the event tree at the bottom.

Someone shortly afterward decided to connect them horizontally on a page instead, and the current bow tie notation was created. The name was given because the result looked somewhat like a man's bow tie. It should be emphasized that only notational differences along with minor changes in content

are involved in these various incarnations of bow tie diagrams. The model being depicted is the same, that is, the chain of events leading to the loss.

Figure 12 shows an example of how the bow tie was originally conceived, with the (now backward) fault tree on the left and the event tree on the right. They are connected by the common hazardous event, which is a light failure in the Figure 12 example.

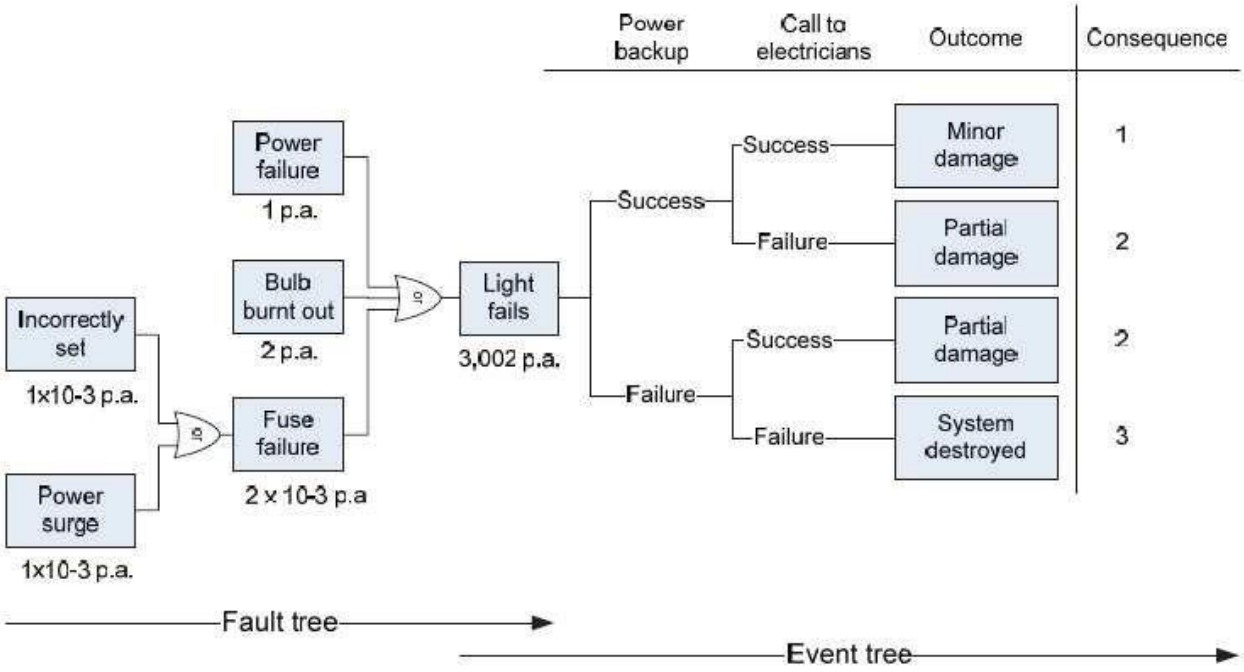


Figure 12: The original conception of the Bow Tie

Bow Tie diagrams are only a way of documenting the analyses generated by forward and backward search techniques such as FMEA, Fault Trees, Event Trees, and HAZOP or other lesser used techniques. Bow Ties are not a new analysis method. The chains of events must still be generated in some way; they are just drawn as a bow tie. Some in the chemical industry, for example, have promoted the use of bow tie diagrams to document the results of HAZOP analyses. Recently in aviation, there seems to be an attempt to bypass the analysis step and just start with the drawing of the bow ties. But without the use of an analysis technique, it is difficult to understand how the information in the bow ties will be generated beyond using very ad hoc, unstructured, and perhaps unreliable approaches such as brainstorming.

Fault trees have a nice graphical notation and bow ties or other graphical depictions are not needed. The other hazard analysis techniques generate so much information for complex systems that they store the results usually in a tabular format. This stored information includes both the chains of events as well as the controls (including but not limited to barriers) designed to try to prevent the events in the chain or to prevent their propagation. The original bow tie diagrams did not include the controls, but these now appear to be commonly inserted into the original bow tie notation.

Figure 13 shows the most common conception of bow tie diagrams today. In order to provide a graphical format, fit them on a page, and include controls or hazard prevention mechanisms as well as mitigation measures, what is chosen to be included must be greatly simplified and the information included must be limited. Without the underlying database of causal information, is a false sense of

security and risk assessment promoted by oversimplifying the causality information provided to the user? Even if there is full documentation of linear causality information in a large database or document, will people ignore it in favor of the simplicity of just using the diagrams? Even worse, are people starting with the bow tie diagram alone and not knowing that they are getting a very distorted view of the risk in the system? The limitations of bow tie diagrams (and in fact, any graphical presentations of causality) are discussed in the next section.

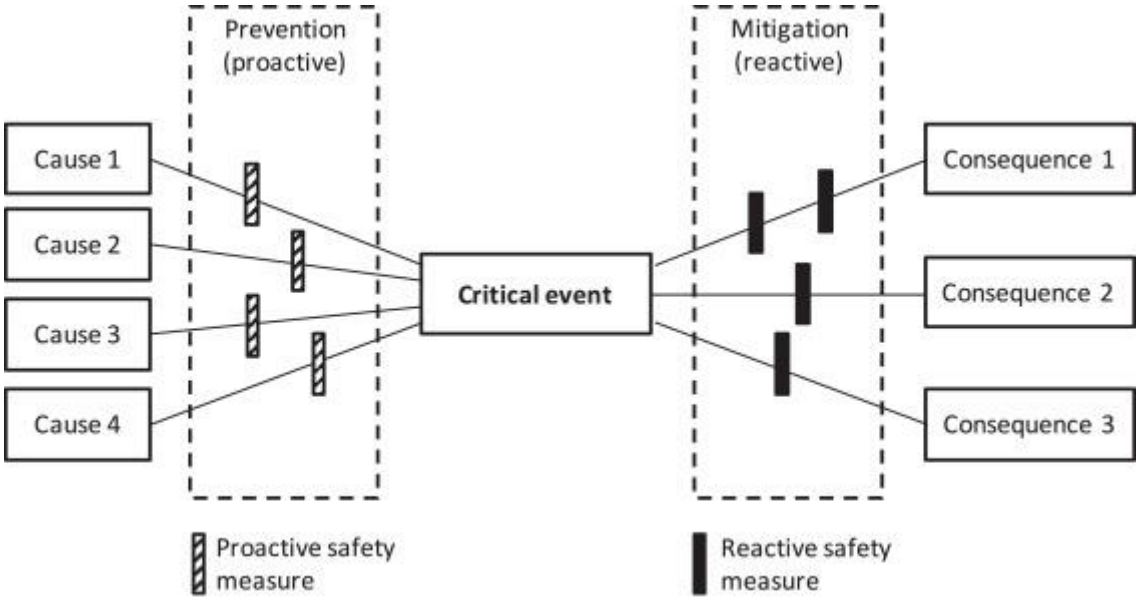


Figure 13: The most common form of Bow Tie diagrams used today.

Limitations of Bow Tie Models

The most important limitation, applicable to all the analysis techniques described so far as well as the bow tie diagram (or similar chain-of-events diagrams), is that they are based on an underlying assumption that accidents are caused by a linear chain of events. They cannot generate nor show the non-linear causality important in accidents today. Therefore, the underlying tables and databases generated by analysis techniques based on assumptions about linear causality in accidents are probably themselves incomplete for complex systems today, no matter how they are displayed. And using a linear graphical notation with boxes and arrows between the boxes ensures that non-linear causality is omitted even if non-linear causality is generated, using modern systemic causality analysis tools, and included in the underlying causality database.

It is tempting to try to find a graphical notation that can help people understand the very large database of results from hazard analysis or even accident analysis. The question really is how to do this and, in the end, whether that goal is practical. Does showing the stored information in a bow tie notation distort the results or limit the ability to use them to investigate accidents or prevent them?

It is clear that the original conception for bow tie diagrams is not practical for complex systems. Putting all the events in the causal chain to the left of the hazard and then all the events after the hazard and up to the actual loss is not possible. Trying to include the controls and even failures of controls (called escalation factors) leading to the hazard as well as the mitigations after the hazard, challenges the practicality even more.

The only way to accomplish this goal is to leave information out and simplify. The bow ties that I have seen recently greatly simplify the causal chains and the mitigation measures to the extent that critical information is excluded. Does this help or hinder safety engineering?

In the case of the most common current Bow Tie notations, what is sacrificed seems to be most of the causal information in order to include, on the same page, information about (some of) the controls to prevent or mitigate the limited causes that are shown (Figure 13). The ANDs are essentially removed from the causal chains (the fault tree) on the left of the hazard, and any ordered recovery actions are omitted. This eliminates all the power of fault trees and event trees to represent detailed causal chains of failure events. What is left is linear chains of events that include only one event. More discussion of this can be found below.

Another change is the inclusion of prevention and recovery actions on the same page. Figure 2 shows these as annotations for the boxes. In complex systems, fault tree and event tree analyses usually document these actions separately, often using tabular formats. There simply is not enough room on a page to include both multi-event chains and the control or recovery actions associated with each event without leaving out almost all the important and useful information.

In order to provide examples, I use here some aviation-related bow tie diagrams from the U.K. CAA webpage. The limitations I point out are not a reflection on those who developed the diagrams but simply a result of them attempting to perform an impossible task. In addition, they are just one conception of bow tie models. Others may exist with slight differences, but the limitations will be the same.

I chose models for the hazard of controlled flight into terrain (CFIT) because recently I have been involved in some CFIT accident analyses and operational hazard analyses and can compare the bow tie results to our results generated using new state of the art modeling and analysis tools. These new tools are not based on the standard chain of failure events model and therefore have the potential to identify many more causes of accidents than just failures. They handle new technology and software, sophisticated human factors, organization design (SMS) and culture, as well as regulatory practices. The new accident causality model is called STAMP and the tools are CAST (used for post-accident analysis of the causes of an accident) and STPA (a proactive hazard analysis tool to assist in preventing accidents before they occur). This paper is not about those tools so only references will be provided for additional information if the reader is interested.[5] These tools are only referenced here to provide information about the usefulness of bow tie diagrams in identifying and preventing the causes of aviation accidents

Figure 15 shows the Bow Tie displaying the event chains, controls, and recovery actions for CFIT in a non-precision approach during IMC (Instrument Meteorological Conditions) or at night. Figure 16 shows CFIT for precision approach in IMC or at night. Figure 17 shows CFIT for arrival or departure in general. In the notation used for the examples in these figures, the blue boxes at the far left represent causes of the hazardous event in the red circle. In all these figures, the hazard is "Terrain separation deteriorating below normal requirements." The boxes between the blue and red circle, white with what looks like a slice from a grey and white solid circle above it so it looks like they are hanging from the line, are the controls used to prevent the failure in the blue box. The controls seem to be unordered in the examples I have seen. On the right side of the red circle containing the hazard (the two sides had to be shown separately to be semi-readable so the right half is shown below the left half of the figure) are controls for mitigating the hazard in the red circle. Finally, the red rounded-corner rectangle at the far right of the diagram shows the loss, which in this case is CFIT resulting in fatalities.

[5] Nancy Leveson*, Engineering a Safer World*, MIT Press, 2012, free pdf version available for free download from the publisher's website. Also, A CAST Handbook and an STPA Handbook are both available for free download from e

Figure 15 had to be split into two pieces to be shown. I tried landscape mode and it still did not fit when shown as a whole. Even split into two, the reader needs to have very good eyesight or a magnifying glass. Of course, tools can be used to magnify figures on the computer screen (you are probably already using one to read Figure 15 if you are reading this paper on-line), but such magnification necessarily reduces the amount of information on the screen at any time.

But readability and human factors concerns are the least of the problems of the two causal chains in Figure 15. I have reproduced the information provided in Figure 15 in a table in order to save the eyesight of the readers of this paper.

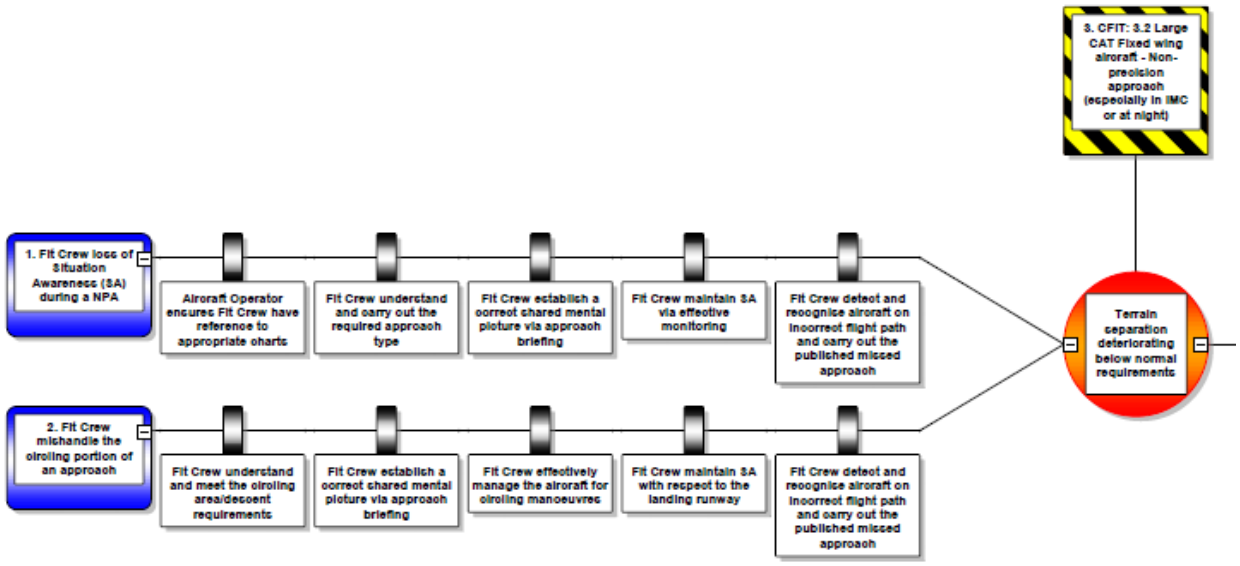


Figure 15: Bow Tie Diagram of CFIT for Non-Precision Approach.

Table 1: Bow-Tie Cause and Consequence Information for a Non-Precision (NP) Approach CFIT (Large Fixed Wing Aircraft)

CAUSE	CONTROLS
FC loss of situation awareness during NP approach	Aircraft Operator ensures FC have reference to appropriate charts
	FC understand and carry out required approach type
	FC establish a correct shared mental picture via approach building
	FC maintain SA via effective monitoring
	FC detect and recognize AC on incorrect flight path and carry out the published missed approach
FC mishandle the circling portion of an	FC understand and meet the circling area/descent

approach	requirements
	FC establish a correct shared mental picture via approach briefing
	FC effectively manage AC for circling maneuvers
	FC maintain effective SA with respect to landing runway
	FC detect and recognize AC on incorrect flight path and carry out the published missed approach

RESULTS IN: Terrain separation deteriorating below normal requirements

Controls after this event (recovery):

- ATCO detects and recognizes incorrect position/altitude and alerts FC
- TAWS alerts FC to inadequate terrain separation
- FC detect and recognize potential conflict visually
- FC carry out terrain avoidance maneuver in response to visual, ATCO or TAWS warning

CONSEQUENCE: CFIT resulting in fatalities

The bow tie diagram represented in Table 1 includes two causes of CFIT during non-precision approach: (1) loss of flight crew situation awareness and (2) the flight crew (FC) mishandling the circling part of the approach. While these causes are true, they are so vague as to not be very useful. Situation awareness is a commonly misused term to mean almost everything under the sun. To be useful, one needs to know exactly what information is missing or wrong in the flight crew's mental model of the situation and why the flight crew is confused about the current situation. For example, what aspects of situation are relevant and have been lost? Why has this occurred? Are the automation displays confusing or providing wrong information, is the flight crew not using the automation, is the information not being provided by the automation or is there no appropriate source (e.g., NOTAMs, ATIS, or charts) where the flight crew can access it or is the information in these places incorrect or missing, is the flight crew too busy or distracted to acquire the information, did they mis-hear the ATCO instructions, etc.? Without considering the actual information lost or why it has been "lost," designed controls to prevent the cause can only be vague, as they are in this case: The first control is for the Aircraft Operator to ensure the FC have reference to the appropriate charts. While this is pretty obvious, it leaves out such questions as *what if the charts are wrong or outdated? What if the FC is distracted and overloaded? etc.*

The rest of the controls or prevention mechanisms essentially say that the accident will be prevented if the FC do everything right. The problem is that accidents occur when someone does something wrong. Telling people not to make mistakes is not very helpful in preventing losses. Compare these "controls" to the engineered controls shown in Figure 1. To be useful in creating better training or procedures, the specification of causes and controls must be much more detailed. The second cause, mishandling circling and its controls is equally vague and useless. Together they are essentially just a definition of CFIT. This is not a criticism of those who made the bow tie diagram. The problem cannot be overcome by simply substituting more specific causes because the problem arises from reducing causal chains into only one box or failure.

One glaring problem with the analysis in Figure 15 is that it implies that all CFIT results from FC errors: it omits not only the details about the FC errors that are needed to address the errors, but also all the other non-FC causal factors of a non-precision CFIT. A group of us recently used CAST to analyze the causal factors involved in the crash of UPS Flight 1354 during a non-precision landing at Birmingham-Shuttleworth International Airport on August 14, 2013.[6] The NTSB report on this accident cited the probable cause as a flight crew "failure."[7] The other identified contributory causes are also associated with the flight crew behavior.

The CAST analysis examines specific behaviors of the flight crew—not something vague like "lost situation awareness"— that contributed to the CFIT as well as explanations for why these behaviors would have seemed like the right thing to do under the circumstances. It is the information about why the behaviors seemed right at the time that are useful in preventing similar accidents in the future. It also pointed out other contributors to the CFIT such as:

- The ATCO[8] did not detect and warn the crew about early descent. He did not know the aircraft was on too low an approach path and did not receive an MSAW (Minimum Safe Altitude Warning). An MSAW (Minimum Safe Altitude Warning) did not alert the ATC to the early descent of the aircraft because the algorithm used was designed such that the aircraft never entered the warning zone.
- The PAPI (a visual system used to help the flight crew line themselves up on the runway in lieu of an instrument landing system, which was not installed on this runway) was designed for height group 3 aircraft while the aircraft involved was a height group 4 aircraft; the PAPI would have been visible for less than 1 second in this case before being obscured by rising terrain. There are no rules limiting the height of aircraft that can land on particular runways as long as the runway is long enough. Without aids such as PAPI or ILS, the approach angle was the only visual cue the pilots had to judge approach angle at night, which is difficult, especially when the runway is surrounded by unlit areas creating what is known as a "black hole" effect.
- The pilots were flying to a runway that did not have an ILS approach as the primary runway was closed due to scheduled one hour maintenance despite the fact that the UPS widebody was scheduled to arrive during that time. It is not known why the maintenance was scheduled at that time but it appears it was due to historical norms and the time set prior to overnight package companies arriving in the early morning hours in large widebody jets.
- The EGPWS (Extended Ground Proximity Warning System) only sounded when the aircraft was already hitting the tops of trees, when not enough time was left to avoid the accident. The deficiencies had to do with the design of the software, UPS not installing software updates that were free but not required by the FAA, and UPS not enabling callouts. The alerts did not escalate as designed due to the proximity of the airport and the terrain on this approach. Using the software version on the aircraft at the time of the accident, the software calculated that the aircraft would be able to safely execute an escape maneuver in the time left. The EGPWS on the aircraft did not contain the latest software enhancements, which were free but needed to be installed. UPS had not done this. UPS assumed that compliance with FAA guidelines was enough to ensure safety.

---

[6] Shem Malmquist, Nancy Leveson, Gus Larard, Jim Perry, and Darren Straker, Increasing Learning from Accidents: A Systems Approach Illustrated by the UPS Flight 1354 CFIT Accident, downloadable from: http://sunnyday.mit.edu/UPS-CAST-Final.pdf

[7] NTSB, *UPS 1354 DCA13MA133.* Retrieved from: https://www.ntsb.gov/investigations/Pages/2014-Birmingham-AL.aspx

[8] Air Traffic Controller

- To trigger ground proximity alerts that would require aggressive enough action from the pilots to avert this particular accident, the software would have had to be enhanced beyond the latest software version, which would be outside the design specifications of the FAA.
- The Flight Management System FMS) "believed" it was below the programmed path due to the actual routing being shorter than the programmed routing, i.e., the FMS assumed it was flying a longer routing so therefore thought it should be at a higher altitude as it was further from the airport. This rendered the FMS vertical path useless. The FMS is not designed to know the routing, but is dependent on pilot actions. The pilot did not sequence the waypoints correctly but was not aware of this fact due to the aircraft tracking the correct course based on ground-based navigational aids (localizer).
- The autopilot will not normally fly past an altitude selected on the MCP. If the minimum descent altitude were entered on the MCP, the autopilot would "capture" that selected altitude. However, the published procedures call for the pilots to instead set the missed approach altitude on their MCP. The missed approach altitude is normally above the final approach fix altitude. As the autopilot now has no constraints on the descent, it will not stop without pilot intervention. No other mechanism aside from selecting an altitude will stop the descent at MDA as the system is designed. It would be possible to create alternative programming to force a disconnect to descend below MDA absent direct pilot action, even allowing for a slight excursion of 50 feet below MDA at intercept. With the current system design, however, the only way to create this feature would limit a full constant angle descent procedure and would not allow for the setting of the missed approach altitude.
- ATCO did not include weather information about variable ceilings in ATIS nor update the weather after ASOS issued a "special observation." He was trying to leave a margin of safety for the pilots as the special observation was an improvement over the previous weather reported on ATIS. It is unclear why the remarks about a variable ceiling were not appended on the ATIS.
- ATCO provided a late descent clearance, putting the aircraft well above a normal descent profile to intercept the final approach course. The aircraft had been held high by Atlanta and then Memphis control due to air traffic control factors.
- The navigation display showed the aircraft on route even though the waypoints had not actually been sequenced, thus misleading the FC into thinking that the waypoints had been sequenced when they had not. At lower range scales, the "extra routing" would not be visible in the VDI or, if visible, would not be salient on the navigation display. Thus, the cues for inadequate programming required interpretation by the flight crew. The design cues were standard for the time the system was designed and continue to be widely utilized by industry. The functionality to conduct profile mode approaches was an addition/modification to an existing system requiring the OEM to work around constraints. The system was created at the request of customers who were, in turn, working to comply with changes in industry practice implemented by regulators. Z
- The pilot flying did not detect that the waypoints were not sequenced, allowing the aircraft to descend below the approach profile by changing modes to one that had too high of a descent angle. He flew when fatigued.
- The pilot monitoring improperly entered the clearance into the FMS, did not make the required callouts, did not challenge the captain's selection of 1,500 feet per minute of vertical speed, and did not properly monitor aircraft position, including altitude. Fatigue and the pace of activity was high at the time. Because the LOC (localizer) approach was used, the aircraft intercepted the course as expected so the lack of proper programming was not apparent. Time compression due to the pace of events and other responsibilities, such as checklists, etc. made the callouts easy to miss. There are a great many callouts and not all help to increase situation awareness.

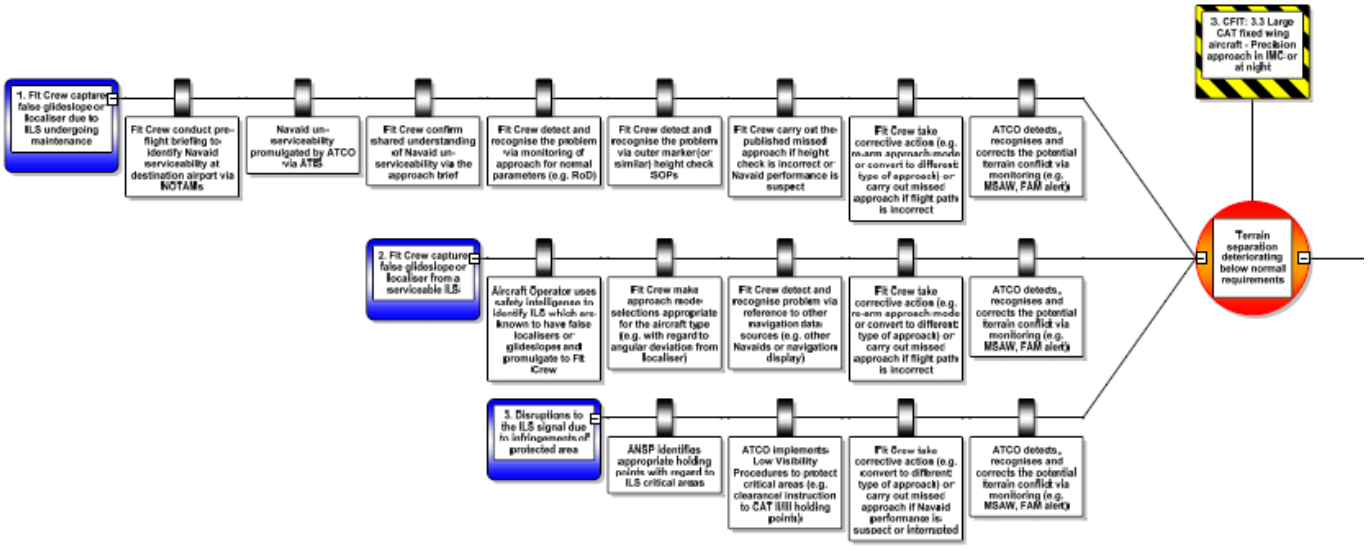
- The design of the aircraft contributed to the confusion on a non-precision approach. It relied on pilot knowledge and procedures. It also had a design where the aircraft would continue to descend below minimum altitudes in a vertical speed approach with the recommended procedures. These procedures are standard in the industry. Also, profile mode was not an initial function in the electronics and required several steps to accomplish. An assumption was made that humans would reliably follow the procedures.
- The weather forecast at BHN indicated low ceilings upon arrival but the dispatcher did not discuss with the flight crew other options for landing on a runway with ILS. The NOTAM (NOTice to Air Men) remarks section that contained the weather information had been removed but neither the dispatcher nor the flight crew were aware of that fact. The only way for the dispatcher to get the remarks would have been to pull them up via a different computer system and that only occurred when a pilot would specifically ask for it. A pilot would not likely ask for this information if they did not know it was missing. UPS had removed the remarks from the weather information provided to pilots either through dispatch paperwork or via weather requests on ACARS. Specifically, UPS requested that UPS's vendor remove weather remarks from the NOTAM to avoid a duplication problem. The information that this was done was not shared with the flight operations department. So, the flight crew did not get accurate weather information from ATIC, the NOTAM, or the dispatcher. The weather remarks were removed from the standard information pilots receive without an adequate assessment of the consequences and without assurance that critical weather information was available to the flight crew.
- The software system used by the dispatcher discouraged him from doing more than just checking that the route was legal.
- The industry trend away from pilots directly interacting with dispatchers (and meteorologists) has led to a reliance on providing data to pilots via printed form, often without discussion or providing context as would occur with direct interaction and discussion.
- UPS did not create effective procedures to mitigate the risk of fatigue (such as briefings and protocols with dispatchers) beyond flight and duty rules. The industry as a whole has not created well-specified protocols nor enforced fatigue standards.
- UPS did not implement procedures that would prevent the autopilot from descending below minimums on a vertical speed approach. Such procedures, however, were not required by Airbus or FAA guidance. UPS did not ensure that the pilots had a complete mental model of how the system performed a profile approach and what requirements needed to be met, and did not enforce a requirement that an approach be immediately abandoned if the aircraft is not stable on the vertical path in the correct mode by the final approach fix.
- The order of the charts provided to the flight crew might have affected the chart selected by the pilot. The accident would not have occurred if they had been flying the RNAV approach as the aircraft would not have tracked the course due to the lack of waypoint sequencing. There was no profile view of terrain on the aeronautical charts to aid the pilot in determining the risk of CFIT during the approach itself.
- There is some suspicion among cargo aircraft pilot associations that cargo aircraft are treated differently by airports, with more risk allowed. Questions have been raised about whether there is more concern for daylight operations than in early morning in darkness with fatigue.

These were not all of the factors found in our UPS BHM CFIT causal analysis, and they are, of course, only the factors involved in this one non-precision approach CFIT. But they are missing from the Bow Tie diagram of CFIT during a non-precision approach shown in Figure 15 and Table 1 and many are missing from the more general CFIT bow diagram in Figure 17 and the Appendix.

The U.K. CAA website suggests that bow-ties diagrams can be used to provide a non-probabilistic risk assessment of the hazard. I totally agree that qualitative assessments are better than than the usual pulling numbers out of thin air. And such a bow-tie model can potentially assist in risk assessment. The problem is that excluding most of the factors that will cause CFIT in practice provides only a false assessment of the risk, which itself can increase the likelihood that such an accident will occur. A very inaccurate risk assessment may be more dangerous than not doing one at all.

I will not go through the bow-tie event chains for CFIT in IMC or at night shown in Figure 17. I have translated the information provided by the bow-tie into a table and the analysis of the problems with it is left as an exercise for the reader. Note, however, that causes related to ILS are included but the FC still seems to be the focus of the causal events. Also, there is again an assumption of single failures and not multiple ones; the latter, however, is more typical in real accidents. Finally, non-ILS causes are omitted. What are the potential non-ILS related causes and possible multiple failures or errors?[9]

Because of the unreadability of Figure 16, I have again translated the information into tabular form. Actually, I don't understand why the Bow Tie format is used as it has severe human factors problems include the difficulty of reading it on paper or on a screen. A simple table is much easier to read in this case.



[9] As a software and automation expert, I am amazed that the aircraft automation doing the wrong thing is not included as a causal factor nor are problems in the pilot-automation interaction.

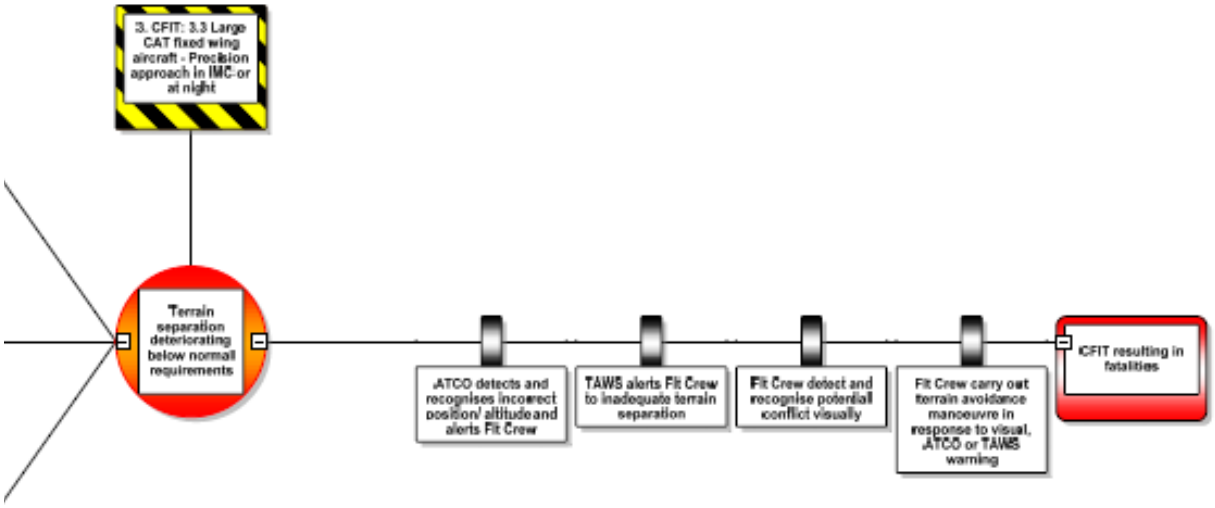


Figure 16: Bow Tie Diagram for CFIT on Precision Approach in IMC or at night

Table 2: Bow Tie Model Information for CFIT when on Precision Approach in IMC or at night (Large Fixed Wing Aircraft)

CAUSE	CONTROLS
FC capture false glideslope or localizer due to ILS undergoing maintenance	FC conduct pre-flight briefing to identify Navaid serviceability at destination airport via NOTAMs
	Navaid unserviceability promulgated by ATCO via ATIS
	FC confirm shared understanding of Navaid unserviceability via the approach brief
	FC detect and recognize the problem via monitoring of approach for normal parameters (e.g., RoD)
	FC detect and recognize the problem via outer marker (or similar) height check SOPs
	FC carry out the published missed approach if height check is incorrect or Navaid performance is suspect
	FC take corrective action (e.g., rearm approach mode or convert to different type of approach) or carry out missed approach if flight path is incorrect
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g., MSAW, FAM alert)
FC capture false glideslope or localizer from a serviceable ILS	Aircraft operator uses safety intelligence to identify ILS that are known to have false localizers or glideslopes and promulgate to FC

	FC make approach mode selections appropriate for the AC type (e.g., with regard to angular deviation from localizer)
	FC detect and recognize problem via reference to other navigation data sources (e.g., other Navaids or navigation display)
	FC take corrective action (e.g., re-arm approach mode or convert to different type of approach) or carry out missed approach if flight path is incorrect
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g., MSAW, FAM alert)
Disruptions to the ILS signal due to infringements of protected area	ANSP identifies appropriate holding points with regard to ILS critical areas
	ATCO implements Low Visibility Procedures to protect critical areas (e.g., clearance/instruction to CAT II/III holding points)
	FC take corrective action (e.g., convert to different type of approach) or carry out missed approach if Navaid performance is suspect or interrupted
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g., MSAW, FAM alert)

RESULTS IN: Terrain separation deteriorating below normal requirements

Controls after this event (recovery):

- ATCO detects and recognizes incorrect position/altitude and alerts FC
- TAWS alerts FC to inadequate terrain separation
- FC detect and recognize potential conflict visually
- FC carry out terrain avoidance maneuver in response to visual, ATCO or TAWS warning

CONSEQUENCE: CFIT resulting in fatalities

To be fair, it is possible that the bow tie diagram for CFIT during non-precision approaches and during IMC and at night that I downloaded from the U.K. CAA website were just two simplified examples and not a real analysis. But the third one for general CFIT seems more complete and about as much as one could realistically show using bow tie diagrams although still substantially incomplete and, because of this, not very useful. The website suggests that it (and the other two) were produced by experts working in groups and seemed to imply a satisfaction with these models.

As can be seen, Figure 17 is unreadable and must be greatly enlarged to be seen, resulting in only small pieces being viewable at any time. I found it unusable, even when I zoomed into the pieces of it (I couldn't see what was connected to what), so I had to once again translate it into tabular form to

understand it. The resulting table is large enough that I decided it was more appropriate to include as an appendix.

Because all CFIT is included in the general model, there are a few more causes, but still only eight single event causes and thus eight event chains with only one box in each chain (italicized words are my additions and are not in the bow tie diagram):

1. ATCO issues incorrect or incomplete clearance/instruction
2. FC misunderstand clearance/instruction
3. FC do not correctly manage AC to achieve or maintain clearance or instruction or correct flight path *(in what ways? why?*)
4. FC mis-set altimeter pressure setting resulting in incorrect actual altitude (e.g., mis-set QNH or low temperature correction (*why?*)
5. Navigation error based on incorrect content within navigation databases or charts and advice provider (*why was it incorrect? Was it originally incorrect? Not updated? Updated incorrectly? the world changed?*)
6. Navigation error due to incorrect FMS entry, incorrect chart selection, or mis-set ground aid by FC
7. Navigation error due to AC position determination (e.g., IRS drift, space weather, miscalibration)
8. Navigation error due to AC position determination (e.g., IRS drift, space weather, miscalibration)

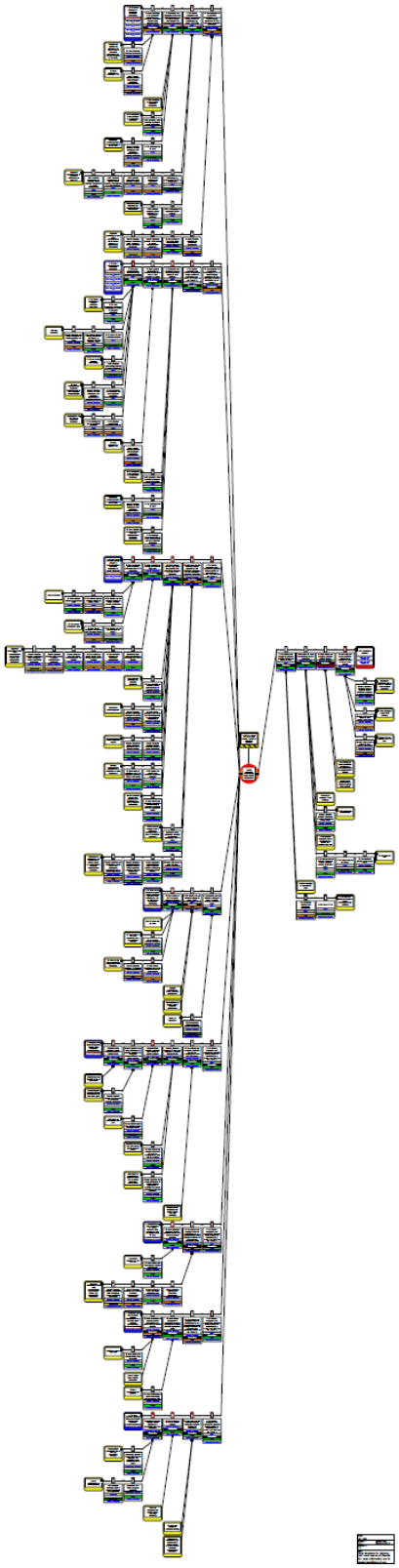


Figure 17: Bow Tie for General CFIT (Arrival or Departure)

These are a small subset of the causes of CFIT and, again, no information is provided about why these errors might occur. For example, why would the ATCO issue an incorrect or incomplete clearance or instruction? The source could be a problem within the controller or within the controller's environment (e.g., the automation being used). Incorrect information could originate upstream of the controller. Why might the FC not correctly manage the automation? Just saying that the FC mismanages the automation without specifying in what ways that could happen does not seem to be useful as it is obvious without providing information needed to design controls to prevent it. The problem could be in the automation design, the FC training, design of the pilot-automation interface, the context including even pressures put on the flight crew through management decisions, weather, the information that the FC was not given or that was given that was wrong or misleading, etc. Why might the content of the navigation databases or charts be incorrect? The information could be incorrect when first generated, become incorrect because it is not updated, or be updated incorrectly. A malicious individual could penetrate the database and change the information. etc., etc. There are a large number of causal factors missing here that are necessary to design effective controls and preventative measures.

This bow tie analysis does include types of (single) failures of the controls, and this is where some of the information that might have been in the causal chains (if more than one event had been included in the causal chain) and I complained about not seeing in the causal "chains" actually is included. For example, for the first cause, ATCO issues incorrect or incomplete clearance/instruction, the "controls" include the FC recognizing the error. But the FC may not recognize it because the potential for conflict is not obvious to the FC at the time of clearance/ instruction or the FC is reluctant to challenge the ATCO. While still too vague and incomplete to be very useful, at least it is a little more complete than the single failure event included as the "cause." The problem is that, with the Bow Tie notation, the failures of controls are spread all over the diagram and therefore difficult to find, and it is probably impossible to identify all the missing ones.

Even if one is generous and tries to find all the potential causes of CFIT somewhere in the Figure 17 bow tie diagram, there are still a tremendous number of missing causes and thus a vast understatement of the risk. In comparison, we used STPA to identify the causes of an unstable approach during landing of a Boeing 777, a major cause of CFIT.[10] We identified 51 unsafe control actions by ATCO, which resulted from 115 scenarios. For the crew, we identified 78 unsafe control actions they could make and 93 scenarios leading to them. We did not include unsafe actions by the automation in this analysis (and neither did the bow tie model in Figure 17), but that would increase the number of scenarios greatly.

As one more example, there is a video[11] by a company selling bow tie drawing tools that purports to provide an overview of "all the different scenarios that could enfold around losing control of a car." There are seven causes identified in their bow tie diagram for losing control of a car:

1. Driver loss of attention (e.g., due to phone, controlling radio, fatigue, eating, etc.)
2. Intoxicated driving
3. Blowout (tire)
4. Unexpected maneuver from nearby car
5. Slippery road conditions
6. Uneven road surface
7. Poor visibility

Note that except for a tire blowout, no failures of the car are in the list. Neither are any of the car automotive controls available on most automobiles today. Without a lot of thought, one could easily

[10] Diogo thesis

[11] The Bow Tie Method in 5 Minutes: https://www.youtube.com/watch?v=P7Z6L7fjsi0

think of more than the seven shown, even those only involving driver actions alone. For example, an animal darts out in front of the car and the driver swerves to avoid it and loses control. More important, all electronics are omitted, as tends to be true in the bow tie diagrams that I have seen. Note that although we don't know exactly the cause of all the unintended acceleration accidents with Toyota automobiles, at the least there is great suspicion that the electronics were involved. Another cause that was hypothesized by Toyota (but did not fit the details of all the cases that occurred) was misplaced floor mats or the driver accidentally depressing the accelerator instead of the brake. None of those are on the list of "all the causal scenarios" listed in the bow tie model above.

My group at MIT works a lot with automobile companies, teaching them how to do complete hazard analyses with sophisticated modern tools. In the process, we have done a lot of examples ourselves. We always include both the simple human errors that usually comprise the causes shown in the bow tie model but also sophisticated problems that are occurring today with automotive electronics, including but not exclusively automated cars. In one case, a graduate student analyzed the causes of the parking assist (automated assistance with parking) function found on many cars today that could lead to an accident (human injury or car damage).[12] Using STPA, she identified over 40 unsafe control actions by the driver that could lead to an accident and about 30 unsafe control actions by the automation for a total of over 70 causes of the hazard. There were hundreds of scenarios that could lead to these mistakes. Clearly the bow tie model mentioned above that is claimed to "provide a clear overview of all the scenarios that could enfold around driving a car"[13] is missing at least 95% of the actual scenarios that could cause real accidents.

In another automotive analysis performed on a real automobile for the manufacturer, Rodrigo Sotomayor analyzed hazards associated with the electric power steering on the car.[14] He found 137 causes of the hazards (engineering design, component failure, manufacturing or maintenance process, interaction, driver error, etc.) and identified 57 prevention actions. Note that this was just the electric power steering.

Another analysis was performed by Major Dan Montes on the flight testing of new military aircraft at Edwards Air Force Base in the U.S.[15] He defined six types of accidents caused by four system hazards. Using STPA, he identified 392 unsafe control actions by the pilots or the aircraft automation that could lead to the hazards.

Blake Abrecht did a causal analysis of a dynamic positioning system on a marine vessel used in oil exploration and extraction.[16] He used the identified causal scenarios to create safety requirements for the design and operation of these systems. He ended up with close to 200 requirements to prevent accidents.

As a final example, Abrecht and several other students (most of whom were Air Force pilots), along with an experienced helicopter pilot, did a hazard analysis for the U.S. Army on the Blackhawk helicopter Caution and Warning system associated with the electrical and fly-by-wire flight control systems.[17] They

---

[12] Megan France, *Engineering for Humans: A New Extension to STPA*, M.S. Thesis, Aeronautics and Astronautics Dept., MIT, June 2017.

[13] The Bow Tie Method in 5 Minutes: https://www.youtube.com/watch?v=P7Z6L7fjsi0video

[14] Rodrigo Sotomayor, *System Theoretic Process Analysis of Electric Power Steering for Automotive Applications*, M.S. Thesis, System Design and Management Program, MIT, June 2015.

[15] Daniel Montes, *Using STPA to Inform Developmental Product Testing*, Ph.D. Dissertation, Dept of Aeronautics and Astronautics, MIT, February 2016.

[16] Blake Abrecht, Systems Theoretic Process Analysis Applied to an Offshore Supply Vessel Dynamic Positioning System, M.S. Thesis, Engineering Systems, MIT, June 2016.

[17] Blake Abrecht, Dave Arterburn, David Horney, John Schneider, Brandon Abel, and Nancy Leveson, A New Approach to Hazard Analysis for Rotorcraft, *American Helicopter Society Development, Affordability, and*

identified 126 unsafe control actions related to the pilot-vehicle interface and the pilot interaction with the helicopter. One example unsafe control action is: "EICAS[18] presents an electrical caution too late for the flight crew to recover the aircraft to a safe condition." Multiple causal scenarios were identified for each of the identified 126 unsafe control actions. Controls were designed to prevent or mitigate them. Hundreds of scenarios were considered and prevented.

Not only are most of these scenarios not included in bow tie models for these types of systems, but including all of them in the bow tie graphical notation would be impossible. Perhaps that is one reason why the bow tie diagrams that people generate all seem to be so incomplete. There also seems to be a tremendous oversimplification of the highly automated systems we operate today. While this may make us feel good about the apparent low risk in our systems and operations, it leads to dangerous complacency and not taking the steps necessary to reduce risk in practice.

In summary, the general CFIT bow tie diagram (shown in the Appendix) includes more causes, controls, and control failures than the first two models reviewed here, but the contents are still very vague/general and incomplete. And there was no way that I could build a mental model of the contents without getting rid of the Bow Tie notation and putting the contents into tables. I have not done a controlled experiment to evaluate the Bow Tie notation, but I have worked with event chain models for almost 40 years and have seen hundreds of examples, so I doubt the problem is strictly within me and not fundamental problems with the design of the notation.

Important questions about any depiction of the results of hazard analysis include: Does the graphical model provide any benefit to justify the cost of producing it? Are the contents so incomplete that the diagram can misdirect attention and create dangerous complacency? Could the inadequacy of the graphical notation itself lead to an accident? These are the questions that need to be answered in evaluating the use of graphical notations to show analysis results.

Conclusions

There are seeming benefits to using linear chain of events models and hazard or accident analysis techniques based on them. The chain of events model is very easy to understand, and it provides a powerful, albeit misplaced, feeling of control by its users: If we can list all the causes of accidents, particularly if there are only a few, then we have the potential to prevent those causes and thus the accidents. In addition, if we stick to hardware failures, there are usually simple engineering solutions (e.g., redundancy and back-up strategies), and we can sometimes come up with probabilities for component failures and easily perform a quantitative risk assessment. But quantitative analysis will only work if component failures are independent and the probabilities are known. And, of course, the model needs to be complete. Almost never are these assumptions true. The most important problem is what the linear analysis methods leave out, which are some of the most important causal factors in accidents today. Using such analysis methods may greatly limit our ability to prevent accidents. [19]

For the complex, software-intensive systems being created and operated today, non-linear accident causality models and analysis techniques are needed. These are not described in this paper, but lots of information about new ones, such as STPA, is available.[20]

Qualification of Complex Systems Specialists' Meeting, Huntsville, Alabama, February 2016.

[18] Engine-Indicating and Crew Alerting System (EICAS), which provides the flight crew with information about the state of the aircraft propulsion system.

[19] Nancy Leveson, Darren Straker, and Shem Malmquist, Updating the Concept of Cause in Accident Investigation, International Society of Air Safety Investigators (ISASI), the Hague, 2019.

[20] Readers might start with Nancy Leveson, *Engineering a Safer World*, MIT Press, 2012 and with the website http://psas.scripts.mit.edu/home/materials

The notations used to document and communicate causal information is an additional problem. The vastly oversimplified accident causation depicted in bow tie diagrams provides management with a very distorted view of the operational risk which, in itself, could lead to accidents. Only a small fraction of the risks in today's systems (e.g., air transportation, railroads, hospitals, chemical plants) will be identified and controlled, leaving companies open to very high levels of risk and liability unless more powerful and complete methods are used.

The problems of creating useful graphical notations are only going to increase if systemic and non-linear causation is considered. While certainly graphical models are nice to have, they just may not be practical or usable for any but the simplest systems or subsystems. More work needs to be done in this area. We are currently experimenting with various types of tabular formats. In some cases, particular types of information may be shown graphically while other information might be better shown in a different format. All information does not need to be documented in the same way.

In the meantime, we need to be careful not to increase risk by using analysis methods and notations to show the results that increase complacency and the potential for accidents by omitting critical information.

Appendix: Tabular Representation of the Information in Figure 17 (Bow-Tie Model of General CFIT)

Bow Tie Tables for General CFIT (Terrain Separation Deteriorating on Arrival or Departure)

CAUSES (left side)

CAUSES	CONTROLS	FAILURES OF CONTROLS	MITIGATIONS
ATCO issues incorrect or incomplete clearance/ instruction	FC challenge ATCO clearance/instruction (e.g., reference to Minimum Sector Radar Vectoring Chart	Potential for conflict not obvious to FC at time of clearance/ instruction (e.g., cannot recognize ATCO error	FC familiarity with local procedures
		FC reluctance to challenge ATCO	ANSP/Aircraft Operator provides joint training to encourage communication
	ATCO detects and recognizes their error while listening to the FC readback (and corrects the clearance/instruction)	ATCO is confident in their original clearance/instruction	
		ATCO inattentive to the readback (e.g., distraction, workload)	ANSP provides "active listening" to ensure proficiency
		Incomplete, heavily accented, or no readback by FC	AC Operator ensures FC compliance with regulatory requirement for language proficiency
			ATCO challenges the FC
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g., MSAW, FAM alert)	Degraded monitoring due to workload, distraction, or other HF.	ANSP workload management via Strategic Sector Capacity Management (e.g., flow management)
			ANSP workload management via Tactical Sector Capacity Management (e.g.,)
			ANSP provides monitoring skills training to ensure monitoring proficiency
			ATCP effective workload

			and distraction management
			Technology (e.g., MSAW) alerts ATCO to change plan
		Degraded system capability due to technical failures or maintenance	ANSP ensures adequate system availability (e.g., maintenance planning
			ANSP contingency for unplanned system outages
	FC maintain SA via effective monitoring (e.g., use of terrain display/ awareness of MSA) and query clearance/instruction	Degraded monitoring or inter-crew communication due human factors (e.g., complacency, distraction)	Aircraft Operator provides monitoring skills training to ensure monitoring proficiency
			Aircraft Operator provides CRM training to ensure communication proficiency
			FC adhere to SOPs that define monitoring roles
			FC effective workload and distraction management
FC misunderstand clearance/ instruction	ATCO issues understandable instruction using standard phraseology	ATCO issues overly long or complex clearance/instruction	ATCO uses progressive clearance/instruction as a form of defensive controlling
		Call-sign confusion	Aircraft operator uses safety intelligence to identify problem call signs
			ANSP utilizes tools to identify problematic call signs (e.g., European Call Sign Similarity Tool)
			ATCO alerts FC to potential problem
		ATCO use of non-standard phraseology	ANSP provides phraseology training to ensure proficiency
		FC misunderstand clearance/instruction	Aircraft Operators ensures FC compliance with regulatory requirement for

			language proficiency
			ATCO recognizes error potential and uses defensive controlling
		Inappropriate use by ATCO of conditional clearance	ATCO minimizes the use of conditional clearance
			ATCO monitors AC post conditional clearance
	FC request clarification of clearance/instruction from ATCO if uncertain of details	FC reluctance to query ATCO	ANSP/Aircraft Operator provides joint training to encourage communication
	ATCO detects and recognizes error or misunderstanding during flight crew readback	ATCO inattentive to the readback (e.g., distraction, workload)	ANSP provides "active listening" training to ensure proficiency
		Incomplete, heavily accented, or no readback by FC	Aircraft Operator ensures FC compliance with regulatory requirement for language proficiency
			ATCO challenges the FC
		FC readback is correct but they have not understood the clearance/instruction	ATCO suspects uncertainty in the FC understanding and uses defensive controlling
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g., MSAW, FAM alert)		
	FC maintains SA via effective monitoring (e.g., use of terrain display, awareness of MSA) and query clearance instruction		
FC do not correctly manage AC to achieve or maintain	FC accurately enter clearance/instruction into FMS/automation	Mis-set altimeter	FC perform altimetry cross-check SOPs
			FC detects mis-set altimeter setting via

clearance or instruction or correct flight path			Mode S
			ATCO detects mis-set altimeter setting via Mode C readout (AC at incorrect level)
		FC errors during data input	FC detect, recognize, and correct error via cross-checks
			ATCO detects mis-set selected level via Mode S
	FC maintain SA, manage and monitor flight plan	Degraded monitoring or inter-crew communication due to human factors (e.g., complacency, distraction)	Aircraft Operator provides 'monitoring skills' training to ensure monitoring proficiency
			Aircraft operator provides CRM training to ensure communication proficiency
			FC adhere to SOPs that define monitoring roles and callouts
			FC adhere to sterile cockpit SOPs during critical phases of flight
			FC effective workload and distraction management
			Automated alerts when approaching or deviating from cleared altitude
	FC manage aircraft to fly within the specified RNP procedure tolerances (e.g., SID, STAR approach)	Poor approach planning (e.g., unstable approach)	FC establish an accurate mental picture via approach briefing
		Automation mismanagement	AC operator provides automation training to ensure proficiency
			FC detect, recognize, and correct the mismanagement of automation via monitoring or automated warnings
		Mismanagement of manual	Suitable automation policy regarding

		flight path control	appropriate use of manual handling
			Aircraft operator provides manual handling training to ensure proficiency
		Adverse environmental conditions (e.g., windshear or wake turbulence)	FC aware of conditions via ATCO (detection systems or PIREPs) and modify approach accordingly
			ATCO applies the correct wake turbulence separation standard
		AC technical defects that destabilize the flight path (e.g, flap or autopilot malfunction)	FC detect and recognize the problem via monitoring or automated warnings and take appropriate action
	FC carry out the published missed approach if RNP procedure tolerances are not maintained (approach scenarios)	FC do not recognize out of tolerances condition due to a loss of situational awareness	Automated below glide slope warnings (on precision approaches) alert FC
		Reluctance to commence a go-around due to human factors (e.g., task fixation, personal/organizational pressure)FC	AC operator has established a safety culture that encourages appropriate use of missed approaches
			AC Operator's simulator training program includes missed approach exercises to encourage proficiency/confidence
			FC approach briefing includes missed approach details to encourage appropriate mental picture
			Effective CRM leads to other FC intervention
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g, MSAW,		

	FAM alert)		
FC mis-set altimeter pressure setting resulting in incorrect actual altitude (e.g., mis-set QNH or low temperature correction	FC adhere to SOP for correct setting of altitude	Both FC mis-set QNH	
		FC forget to set QNH (especially with a low transition altitude)	FC detect the omission during appropriate checklist
		FC do not make the appropriate Temperature Error Correction	Aircraft Operator provides TEC training to ensure proficiency
			FC detect, recognize, and correct calculation error via cross-checks
	ATCO detects mis-set altimeter setting via Mode S	Variable availability of Mode S downlink capability	
		Not mandatory to operate Mode S downlink performance	
	ATCO detects mis-set altimeter setting via Mode C readout (AC at incorrect level)	Mode C is inoperative	Dual transponder carriage provides system redundancy
Navigation error based on incorrect content within navigation databases or charts and advice provider	Database/Chart providers quality assurance process to identify errors	Coding errors not obvious at time of publication	
	Aircraft Operator uses safety intelligence to identify anomalies for navigation databases/ charts and advise provider	Processing time (between problem identification and next chart cycle)	Aircraft Operator informs FC and/or establishes alternative procedures
	Database/chart provider maintains data validity via update program	MRO does not load updates into FMS	FC detect out of date database during pre-flight checks
	Aircraft Operator ensures FC have reference to current database/chart	Mismatch between FMS data and on-board approach charts	FC detect and recognize mismatch via SOP for comparison of chart and FMS display
		Poor clarity or correctness of approach charts (AIP or during third party modifications)	Aircraft Operator uses safety intelligence to identify problem approaches to feedback for special procedures

			(and chart provider)
	FC detect and recognize mismatch via SOP for comparison of chart and FMS display	Comparison not completed due to human factors (e.g., high workload)	
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g., MSAW, FAM alert)		
Navigation error due to incorrect FMS entry, incorrect chart selection, or mis-set ground aid by FC	FC approach/ departure briefing confirms correct FMS, chart, and Navaid selection	Last minute changes, e.g., SID or STAR	FC update plan according to changed requirements
	FC detect and recognize error via maintaining SA (e.g., by monitoring terrain/display/ awareness of MSA) and take appropriate action	Degraded monitoring or inter-crew communication due to human factors (e.g., complacency, distraction)	Aircraft Operator provides "monitoring skills" training to ensure monitoring proficiency
			Aircraft Operator provides CRM training to ensure communication proficiency
			FC adheres to SOPs that define monitoring roles
			FC effective workload and distraction management
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g., MSAW, FAM alert)		
Navigation error due to AC position determination (e.g., IRS drift, space weather, miscalibration)	FC awareness of potential system degradation (e.g., RAIM/ NOTAM/ aircraft detects	Inability to access RAIM data	FC obtain data during briefing state or in-flight
		Lack of usable space weather information	
	FC receive downgrade of navigation accuracy warning from aircraft and take appropriate action	FC misdiagnose the problem	FC proficiency in system knowledge and associated SOPs

	FC detect and recognize error via maintaining SA (e.g., by monitoring/ terrain display/ awareness of MSA) and take appropriate action		
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (e.g, MSAW, FAM alert)		
FC continue the approach below MDA/DH without visual reference	Aerodrome operator ensures runway lighting is in accordance with ICAO standards	Runway lights unavailable due to WIP (e.g., resurfacing runway)	Aerodrome Operator issues NOTAMs that highlights WIP (available to and understood by FC
		Lack of maintenance on runway lighting	NAA conducts oversight audit to ensure compliance with regulations
			Aerodrome SMS audits to ensure compliance
	FC adheres to SOPs for MDA/DH callouts	SOPs not completed due to human factors (e.g., high workload)	
	FC carry out published missed approach if required visibility not maintained	FC task fixated on continued landing	
		Reluctance to commence a GA due to human factors (e.g., task fixation, personal/organizational pressure	
	ATCO detects, recognizes, and corrects the potential terrain conflict via monitoring (MSAW, FAM alert)		

Bow Tie Table for Consequences (Right Hand Side): CFIT Resulting in Fatalities

HAZARD: Terrain separation deteriorating below normal requirements.

MITIGATION	CONTROLS	FAILURES OF CONTROLS/ MITIGATIONS
ATCO detects and recognizes incorrect position/altitude and alerts FC		Limited coverage of surveillance radar
	ATCO effective workload and distraction management	ATCO does not detect the problem due to distractions/ high workload
	MSAW alerts ATCO	
TAWS alerts FC to inadequate terrain separation		Incorrect or outdated TAWS database
	Aircraft Operator prohibits operations to critical aerodrome as per MELs	TAWs temporarily unserviceable
		TAWS position inaccurate due to substandard or absent GPS position data feed
	ATCO detects mis-set altimeter setting via Mode C readout (AC at incorrect level)	FC mis-set QNH
	ATCO detects mis-set via Mode S	
	FC perform altimetry cross-check SOPs	
FC detect and recognize potential conflict visually		Low visibility (or night) conditions resulting in a limited opportunity to respond before collision
		AC lacks performance to achieve required climb gradient
FC carry out terrain avoidance maneuver in response to visual, ATCO, or TAWS warning	Aircraft Operator's simulation program includes terrain avoidance exercises to ensure proficiency	Inadequate proficiency due to a lack of exposure to the required responses
	Effective CRM leads to other FC member intervention	FC disregard a valid TAWS warning
	Effective CRM leads to other FC member intervention	Delayed FC response to ATCO alert